

[illegible]

2. US-08-223-263-2 (1-1795)
Adrenalin receptor subtype beta-1 gene.

[illegible][illegible]

[illegible]

CCCTGGCCCCAGGAGCAGCTGATGAGACACATTGGTAAGTGTCCCTTTCAAGTGCCTTTCCATTAAAGACC	1180	1790	1800	1810	1820	1830	1840
AGTA--TAGCTCTTCCCTCTTCCACACCACTTGGCCACCCACCTGTGTGACAGTCCACCCCTGCTTCTCTGA	1130	1140	1150	1160	1170	1180	1190
AGAAAGGAGCGCTTTGAC---TGCAGCC---TGTGGGTGGCTC-GGTACGC---AGGAATGAT--GGA	1850	1860	1870	1880	1890	1900	
CCCTTCTGCTCCAGGCCACCC--CTACAGACCT-----CTTCTAAACATCATCTACACCACCTCCAGAA	1200	1210	1220	1230	1240	1250	
CGTTTAGCAGAGTGGGCTTGTCCATCTTTGAGCTTTTGGGCTTCAAGATCA-GCTGGAAGAGTCTACCGCG	1910	1920	1930	1940	1950	1960	1970
TCATCTCTCAGAAAGGTAAAGTGTCTCAGAC-TGCCGACATAGCATT---GTCTCATG-TACAGCTCCCG	1260	1270	1280	1290	1300	1310	1320
ACGATTTCA--AAG--AAG--TCTTACCACATCTG-TGATATTTCTTCTCGTGGCGCTGATTAACAACCGC	1980	1990	2000	2010	2020	2030	
TTCCCTCGAGGGCGCCCTTGAGCAACACTGCAGACATTTCTCTACTTT---CTCTGAAACCCAAAGCCCT	1330	1340	1350	1360	1370	1380	1390
TGACCAAAATG---ACT---TACAAAGAGAAAGCTGGCTGTTTAAAGGTTCCACAGGTGAGACACA	2040	2050	2060	2070	2080	2090	2100
TGGTA--AAAGGATACACA--GGACTGA--AAAGGAATCAATTTTCACTGTACATATTAAACCTTCAGAACGC	1400	1410	1420	1430	1440	1450	1460
TGGCAGCCGCGGACACATGGCAGTGGAGGACAGAAAGC--TGAGAGCT-CACA-TCTCAACCAAAAGTTGA	2110	2120	2130	2140	2150	2160	
TATTTTTTTAAG--CTATCAGCAATATCTATCAGACAGCAGCTCTTGTCTATTTTCTCAGAAATTTGCG	1470	1480	1490	1500	1510	1520	1530
GTGACACTGGAAGACTATC--CCCTCCCC--CAGCCCACTCAGCAAGGCTCCA--CCCCCTCTAAGAGTTTC	2170	2180	2190	2200	2210	2220	2230
CAACTCAGCTATTTCTATCATCTCTTTTCTGTGA---TAACCTTCGAAAGGCGCTGGGCTG	1540	1550	1560	1570	X		
CA--TGGCT--CCCTAACA-GCTGGCGCAATAGACCAAGCTTCAAT	2240	2250	2260	2270	2280		

4. US-08-223-263-2 (1-1795)
003684 PAT744 clone of gene encoding activated T cell mRNA

ID	003684 standard; DNA; 659 BP.
AC	003684/;
DT	12-SEP-1990 (first entry)
DE	PAT744 clone of gene encoding activated T cell mRNAs.
KM	Peripheral blood mononuclear cell; PBMC; lymphokine;
KW	cytokine; mitogen; ds.
OS	Homo sapiens.
FH	Key
FT	Location/Qualifiers
CDS	74..349
FT	/tag= a

PN US7312001-A.
PD 13-MAR-1990.
PR 16-DEC-1988; 312001.
PR 16-DEC-1988; US-312001.
PA (USSH) US Dept Health & Human.
PI Siebenlist U, Leonard WJ, Zipfel PJ, Irving SG, Kelly K;
DR WPI; 90-139708/18.
DR P-PSDB; R04222.
PT New lymphokine-cytokine genes -
PT Isolated using mRNA from activated human peripheral blood
PT mononuclear cells and T cells.
PS Disclosure; 184pp; English.
CC The lymphokine/cytokine-like proteins are associated with the
CC inflammatory response and/or have mitogenic activities. Antigens
CC raised to the proteins may be useful in detection and purification,
CC especially in bioassays of various tumour cells or genetic defects
CC in the inflammatory response.
CC The best copy of the sequence available is still unclear and some
CC errors in the sequence may occur.
SQ Sequence 659 BP; 161 A; 181 C; 124 G; 193 T;

Initial Score = 133 Optimized Score = 251 Significance = 7.46
Residue Identity = 47% Matches = 302 Mismatches = 268
Gaps = 68 Conservative Substitutions = 0

X
TCTTCTAC---CAT-CTGCTCCCGAGAGGCTG-CCTG-CGTGCACTTGGCTCGA
AGMAACTCTTTTCCACATATCATAGAGCTCTC-GTG-ACGTCTCTCTCTCTCATG---CTAGTA
50 60 70 80 90 100 110 120
GC--CCTTCCACCCGATATGATTCCTCACCCTTG--CCGCGCTTGGCCCGCCACTT-CTGCCGCA-
110 120 130 140 150 160 170 180
GCTGCTTCTGCTCTAGCACTCTGACACCAATGAGGCTGACCCCTCCACCGCTGCTCTTTTAC
120 130 140 150 160 170 180
130 140 150 160 170 180 190
AGTCAAGAGCCCTAAGCCGCTCCATGCGCCCGAGAGAGATTCAGGAGGCGCCCAACAGAGGAGCAG
110 120 130 140 150 160 170 180 190
ACCGCAG---GAAGCTTCTGCGAATCTTGTGTGA-GATTACTATAG-CAAGAGCCTGCTGCCAG
190 200 210 220 230 240 250
CCAGCTGTGTATCCAAACAAAGAGCAAGCTGTGCTGATCCAGTGAATCTGTGGTC-CAAGA
260 270 280 290 300 310 320
CCAGC-CAGACA-CCCCGGCA---GAATGAGC-TGACTGAATTCCTCTCCGTC-CTCATGCTTCTCTTA
110 120 130 140 150 160 170 180 190
CCAGCTGTGTATCCAAACAAAGAGCAAGCTGTGCTGATCCAGTGAATCTGTGGTC-CAAGA
260 270 280 290 300 310 320
CTCAGAGCCTAAGCTGTGCTCAGCCCGGCTCTCTGCT-TGTGACCTCCAGCTCTCAGTAACTCTTCTGCT
110 120 130 140 150 160 170 180 190
GTACGAGTATGAC-CTG--GAAGTAACT-GAGCTGCTCAGAGA-CAAGAACTTCAAGGAA-GGTCACT
330 340 350 360 370 380 390
330 340 350 360 370 380 390 400
GATCCCATCTCTTC-AAGCAGACTGAGCAGTCCCAAGGCTTCCCTTGCCTAACAACCTGCTGCT
110 120 130 140 150 160 170 180 190
GA-GCCCCGATGCTTCTCATGACACACTTCCT--CCATA--CTCAGAGCTCTCTCCCGCAGTCTG-T
400 410 420 430 440 450

410 420 430 440 450 460
GCTGCTGAGACTTACCTTGGAGATGAGAAACCCAGTGGAGA-GACCAGGACAGCA--CAATCT
110 120 130 140 150 160
CCCTTCTCTTAATTAACTT--TTTATGTCGCGCTGTATTGATAGGTCTCATTTCCATTATTTAAT
460 470 480 490 500 510 520
470 480 490 500 510 520
GG--GAGC--AGTGACCCCTTCTGCTGAGAGGAGATGAGGAGCAGCGGAGCAGCACTTGGCTCT
110 120 130 140 150 160 170 180 190
AGTTTGCCAAAGATTAATCTCCCTATG--GGATGG--TCACCTG--TCACCTTCTCTCCCTTGCA
530 540 550 560 570 580 590
540 550 560 570 580 590 600
CATCCCTCTG--GGCAGCTTCTGAGCAGGCTGCTCTCTCTTGGGGCCCTGAGAGCCTCTGAGAC
110 120 130 140 150 160 170 180 190
AAT--ACATGATTAACATTTG-ATTCTG-TGTGTTTCAATTAATACTTTAAAAA
600 610 620 630 640 650
X

610
CCAGCTTCTCT
5. US-08-223-263-2 (1-1795)
Q05748 EPO receptor gene.

ID 005748 standard; cDNA; 1883 BP.
AC 005748;
DT 04-JAN-1991 (first entry)
DE EPO receptor gene.
KW Erythropoietin; Diamond Blackfan anaemia; polycythemia vera; ss.
OS Homo sapiens.
FH Key
FT CDS Location/Qualifiers
FT CDS 145..1759
PN WO9008822-A.
PD 09-AUG-1990.
PF 01-FEB-1990; U00635.
PR 03-FEB-1989; US-306503.
PA (GENE-) GENETICS INST INC.
PA (WHIT-) WHITEHEAD INST.
PI D'andrea A, Wong G;
DR WPI; 90-260931/34.
DR P-PSDB; R06512.
PT Erythropoietin receptor and gene - used for developing reagents
PT and systems to control and study erythropoiesis.
PS Claim 7; Fig 2; 53pp; English.
CC The sequence was obtd. from a clone isolated from a commercially
CC available human genomic cDNA library in phage Lambda fix
CC (Stratagene). The sequence encodes a type I transmembrane protein
CC with binding affinity for EPO. The gene and recombinant EPO
CC receptor produced on expression of the DNA are used to develop
CC receptors and systems to control and study erythropoiesis. It is
CC believed that the EPO receptor is dysfunctional in individuals
CC with Diamond Blackfan anaemia, and may be hyperactive in poly-
CC cythemia vera.
CC See also Q05747 (murine EPO receptor clone).
SQ Sequence 1883 BP; 320 A; 607 C; 564 G; 372 T;
Initial Score = 133 Optimized Score = 653 Significance = 7.46

1270 1280 1290 1300 1310 1320
CNCAGGAGGTAAAGTCTCAGA----CACTGCCGCATACAGCAT--TGCTCATGTAAAGC--TCCCT-T
||||| | ||||| | ||||| | | | ||| |

[illegible]

11. US-08-223-263-2 (1-1795)
Q06470 Sequence encoding complete form of human RF-X prot

ID Q06470 standard; DNA; 3086 BP.
AC Q06470;
DE 25-FEB-1991 (first entry)
DT Sequence encoding complete form of human RF-X protein.
KW MHC class II; down regulation; autoimmune disease; ds.
OS Homo sapiens.
FH Key
FT CDS Location/Qualifiers
FT /*tag= a
PN MOJ012812-A.
PD 01-NOV-1990.
PF 18-APR-1990; E00625.
PR 18-APR-1989; EP-106944.
PR 14-MG-1989; EP-115008.
PA (MACH/) Mach B.
PI Mach B;
DR WPI; 90-348429/46.
P-PSDB; R08338.
PT Purified proteins and compsns. - regulate expression of MHC class II genes and bind to controlling DNA sequences.
PS Claim 11; Fig 9; 83pp; English.
CC The sequence encodes a protein which regulates the expression of MHC class II genes by binding to DNA sequences which control this expression. The protein causes down-regulation, useful for the prevention and treatment of autoimmune diseases such as Insulin Dependent diabetes, Multiple Sclerosis, lupus Erythematosis and Rheumatoid Arthritis. The protein can also be used for screening CC and identifying substances capable of inhibiting the expression of the MHC II genes.
CC See also Q06469 and Q07000.
SQ Sequence 3086 BP; 603 A; 1109 C; 934 G; 440 T;

Initial Score = 128 Optimized Score = 756 Significance = 7.13
Residue Identity = 48% Matches = 914 Mismatches = 766
Gaps = 223 Conservative Substitutions = 0

X 10 20 30 40 50 60 70
TCTTCACCAACATCTGTCCAGAGGGCTCGTGCTGTGA--CTTGGGTCTGGAGACCCTTCACCCGGG
| | | | | | | | | | | | | | | | | | | |
GCCA-AAAGCCTTTCCGACAGACAAG--TGGGAGAAGGGAGAAATTAAAAAAAAGCCTTT
X 10 20 30 40 50 60
ATAGATTCCCAACCCCTGGCCGCCTTTGCCCAACCTTACTC-TGCCAAGA-AGTCAAGAAGCTTAAGCGCG
| | | | | | | | | | | | | | | | | | | |
ATTATT-ATCA--TTTCCCACCGCTTG-GCATGGCAACAGCGCTTAATCTGACTACAGCACGCCCGCG
70 80 90 100 110 120 130 140
150 160 170 180 190 200
CTCCATGGCCCCAGAAAGATTTCAGGGAGAGGCCCCCAAAAGGAGACCA--CGCCA--GCCAGAACCCCG
| | | | | | | | | | | | | | | | | | | |
CACCAT---CCGACG--CGCCACA--GGCCCCGCCCAAGCCAGCCAGCCGACCGCCACCGCCACACCCCGCA
140 150 160 170 180 190
GCCAAGATGAGCTGACCTAAATTTGCTCTCTGTGTCATGCTTCTCT-AACTGCAA-GGCTAAC---GCTGT

-----TCTTTATGGCCGTCGAACCGGCC--GTCTGGGCACCAGAGCACA CTCCAAGTACACTCTTAATGGCGCTTG
1600 1610 1620

1700 1710 1720 1730 1740 1750 1760

TATCATCTTTCAGTCGG--GACCTGTATCCCATATTCTTAAACAGATCTTACTCTTGGAATAAGTGAAGCTTTT
+ ||| | | | | | | | | | | | | | | | | | | |
T-TC--TCCGACAGAAGAGGCGCTCAAAGC-CATCC-AAGAGAT-GGAAGGACATGACAAAGCGCGTGGCGGGT
1700 1710 1720 1730 1740 1750

1770 1780 1790 X
CTCTCAGAAAAAAA-----AAAAAAAAA
||| | | | | | | | | | | | | | | | | | |
GGGCGACAGCCGACGACGCGCTGTCCGACATCACGCGCCCCAGGT
1760 1770 1780 1790 X 1800

12. US-08-223-263-2 (1-1795)
Q24676 Encodes partial MPVU-env related polypeptide.

ID Q24676 standard; DNA; 552 BP.
DC Q24676;
AD 05-NOV-1992 (first entry)
DE Encodes partial MPVU-env related polypeptide.
KM Myeloproliferative leukaemia virus; envelope protein; gp70;
KW haematopoietic cell; Growth Factor receptor; ss.
OS Myeloproliferative leukaemia virus.
FH Key Location/Qualifiers
FT CDS 1..552
FT /*tag= A
PN MO9207074-A.
PD 30-APR-1992.
PF 19-OCT-1990; F00762.
PR 19-OCT-1990; WO-F07623.
PA (INRNA) INSEEM INST NAT SANTE & RECH MEDICALE.
PI Chambouin M, Gisselbrecht S, Peniolelli JF, Souyri M;
P1 Tambourin P, Varlet P, Vigon I, Wendling F;
DR MPI; 92-167154/20.
DR P-PSDB; R27660.
PT Polypeptides similar to v-mpl protein of MPVU - for diagnosis and treatment of myeloproliferative diseases
PS Claim 10; Page 57-8; 75pp; French.
CC This nucleotide sequence is one of 6 claimed sequences which code for polypeptides related to MPVU env protein. The proteins have one or more of the following properties: they are involved in the ligand-fixing or signal-transmitting function of haematopoietic growth factor receptors; they are recognised by antibodies to the CC protein with GENSEQ accession number R23970; when produced from CC the MPVU genome they can induce/promote proliferation of CC haematopoietic cells and/or are involved in differentiation of CC haematopoietic cells. The precise identity of sequence Q24676 is CC not further described in the specification.
CC See also R23971 and Q24674-7.
SQ Sequence 552 BP; 106 A%; 193 C; 135 G; 118 T;

[illegible]

DE Linked Kallmann Syndrome (KS) .
 KM X-linked Kallmann syndrome; ADMIX gene; diagnosis; PCR; ss.
 OS Homo sapiens. Location/Qualifiers
 FH CDS
 FT /tag= a 151..357
 FT exon
 FT /tag= b 358..405
 FT exon
 FT /tag= c 406..468
 FT exon
 FT /tag= d 469..691
 FT exon
 FT /tag= e 692..876
 FT exon
 FT /tag= f 877..1006
 FT exon
 FT /tag= g 1007..1212
 FT exon
 FT /tag= h 1213..1357
 FT exon
 FT /tag= i 1358..1504
 FT exon
 FT /tag= j 1505..1599
 FT exon
 FT /tag= k 1600..1771
 FT exon
 FT /tag= l 1772..1992
 FT exon
 FT /tag= m 1993..2134
 FT exon
 FT /tag= n 2134..2193
 FT exon
 FT /tag= o
 FT /tag= 14
 PN MO307267-A.
 PD 15-APR-1993.
 PF 09-OCT-1992; F00956.
 PR (INSP) INST PASTEUR.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Claverie J, Hardelin J, Legouis R, Leuilliers J,
 PI Intalla G, Petit C;
 PI WPI 93-134456/16.
 DR P-PSDB; R34445.
 PT Nucleic acid sequence of gene with X-linked Kallmann syndrome -
 useful for diagnosing Kallmann syndrome by amplification to

PT detect genetic anomalies
 PS Claim 1; Fig 6; 60pp; French.
 CC The nucleic acid sequence is derived from the ADMIX gene associated
 CC with KS (or Hypogonadotropic hypogonadism and anosmia).
 CC Oligonucleotide pairs which act as primers for specific amplification
 CC of the gene are used in amplification methods to detect genetic
 CC anomalies which cause KS. The primer pairs corresp. to the coding
 CC and non-coding regions of exon 1 of the ADMIX gene and one pair each
 CC for the other 13 exons.
 SQ Sequence 6314 Bp; 1884 A; 1274 C; 1246 G; 1910 T;
 Initial Score = 120 Optimized Score = 751 Significance = 6.60
 Residue Identity = 47% Matches = 922 Mismatches = 748
 Gaps = 251 Conservative Substitutions = 0
 X 10 20 30 40 50
 60 CTCTCCACCCGAGATGATTCCTCACCCTTGCCGCCCTTACTCTG-CCCAAGATGACAG
 120 ATT-GCAAGCC-ATAACGTAAGTGTGAGTGTGAGAGAACTCCAGCCAGACTGTG-AC-TATGTTGTGA-
 1270 1280 1290 1300 1310 1320 1330
 130 AGCCTAAGCCGCTCCATGAGCCGACGAGAGAGATTCAGAGGAGAGCCCAAGAGGAGAGCCAGCAG
 1340 1350 1360 1370 1380 1390 1400
 ATGACACATGCAACCAACAAAGACAGCTT-GTGAA-ACAGAAAGGTGAATTCAAACACA
 1410 1420 1430 1440 1450 1460 1470
 ACACCCGCGCAGATGAGTGAATGCTC-CTGTGTGTC-ATGCT-TCCTCACTGCAAGGCTAA
 1480 1490 1500 1510 1520 1530
 CGCTGTCAAGCCGCTCTCTCTCTG-TCACCTCCAGTCTCACTGTAATGCT-TCGTGATCCCATG-
 1540 1550 1560 1570 1580 1590 1600
 --CTG-CAAGTTAAAGT-CTACTGCAAGAGAGAGAGATCCCACTGTCAACCATATCATG--TCGGGTG
 1610 1620 1630 1640 1650 1660 1670
 GTCCTGCTGCTCTCTGAGCTTGAAGTTGGAGAGATGAAACCCAGATGAGAGACCA--GGC--AC
 1680 1690 1700 1710 1720 1730 1740
 ATGATCTTCAGATCT-CTCATTTCTCCAGAGAT--AA--GGTACCTGTCA-ACCAATACGCCAAA
 1750 1760 1770 1780 1790 1800 1810
 AGACATTC--TCGAGCAG-TG-ACCCTT-CTCTGAGAGAGATGAGACAGAGGAGACACTGGAC
 1820 1830 1840 1850 1860 1870 1880
 1890 1900 1910 1920 1930 1940 1950
 1960 1970 1980 1990 2000 2010 2020
 2030 2040 2050 2060 2070 2080 2090
 2100 2110 2120 2130 2140 2150 2160
 2170 2180 2190 2200 2210 2220 2230
 2240 2250 2260 2270 2280 2290 2300
 2310 2320 2330 2340 2350 2360 2370
 2380 2390 2400 2410 2420 2430 2440
 2450 2460 2470 2480 2490 2500 2510
 2520 2530 2540 2550 2560 2570 2580
 2590 2600 2610 2620 2630 2640 2650
 2660 2670 2680 2690 2700 2710 2720
 2730 2740 2750 2760 2770 2780 2790
 2800 2810 2820 2830 2840 2850 2860
 2870 2880 2890 2900 2910 2920 2930
 2940 2950 2960 2970 2980 2990 3000
 3010 3020 3030 3040 3050 3060 3070
 3080 3090 3100 3110 3120 3130 3140
 3150 3160 3170 3180 3190 3200 3210
 3220 3230 3240 3250 3260 3270 3280
 3290 3300 3310 3320 3330 3340 3350
 3360 3370 3380 3390 3400 3410 3420
 3430 3440 3450 3460 3470 3480 3490
 3500 3510 3520 3530 3540 3550 3560
 3570 3580 3590 3600 3610 3620 3630
 3640 3650 3660 3670 3680 3690 3700
 3710 3720 3730 3740 3750 3760 3770
 3780 3790 3800 3810 3820 3830 3840
 3850 3860 3870 3880 3890 3900 3910
 3920 3930 3940 3950 3960 3970 3980
 3990 4000 4010 4020 4030 4040 4050
 4060 4070 4080 4090 4100 4110 4120
 4130 4140 4150 4160 4170 4180 4190
 4200 4210 4220 4230 4240 4250 4260
 4270 4280 4290 4300 4310 4320 4330
 4340 4350 4360 4370 4380 4390 4400
 4410 4420 4430 4440 4450 4460 4470
 4480 4490 4500 4510 4520 4530 4540
 4550 4560 4570 4580 4590 4600 4610
 4620 4630 4640 4650 4660 4670 4680
 4690 4700 4710 4720 4730 4740 4750
 4760 4770 4780 4790 4800 4810 4820
 4830 4840 4850 4860 4870 4880 4890
 4900 4910 4920 4930 4940 4950 4960
 4970 4980 4990 5000 5010 5020 5030
 5040 5050 5060 5070 5080 5090 5100
 5110 5120 5130 5140 5150 5160 5170
 5180 5190 5200 5210 5220 5230 5240
 5250 5260 5270 5280 5290 5300 5310
 5320 5330 5340 5350 5360 5370 5380
 5390 5400 5410 5420 5430 5440 5450
 5460 5470 5480 5490 5500 5510 5520
 5530 5540 5550 5560 5570 5580 5590
 5600 5610 5620 5630 5640 5650 5660
 5670 5680 5690 5700 5710 5720 5730
 5740 5750 5760 5770 5780 5790 5800
 5810 5820 5830 5840 5850 5860 5870
 5880 5890 5900 5910 5920 5930 5940
 5950 5960 5970 5980 5990 6000 6010
 6020 6030 6040 6050 6060 6070 6080
 6090 6100 6110 6120 6130 6140 6150
 6160 6170 6180 6190 6200 6210 6220
 6230 6240 6250 6260 6270 6280 6290
 6300 6310 6320 6330 6340 6350 6360
 6370 6380 6390 6400 6410 6420 6430
 6440 6450 6460 6470 6480 6490 6500
 6510 6520 6530 6540 6550 6560 6570
 6580 6590 6600 6610 6620 6630 6640
 6650 6660 6670 6680 6690 6700 6710
 6720 6730 6740 6750 6760 6770 6780
 6790 6800 6810 6820 6830 6840 6850
 6860 6870 6880 6890 6900 6910 6920
 6930 6940 6950 6960 6970 6980 6990
 7000 7010 7020 7030 7040 7050 7060
 7070 7080 7090 7100 7110 7120 7130
 7140 7150 7160 7170 7180 7190 7200
 7210 7220 7230 7240 7250 7260 7270
 7280 7290 7300 7310 7320 7330 7340
 7350 7360 7370 7380 7390 7400 7410
 7420 7430 7440 7450 7460 7470 7480
 7490 7500 7510 7520 7530 7540 7550
 7560 7570 7580 7590 7600 7610 7620
 7630 7640 7650 7660 7670 7680 7690
 7700 7710 7720 7730 7740 7750 7760
 7770 7780 7790 7800 7810 7820 7830
 7840 7850 7860 7870 7880 7890 7900
 7910 7920 7930 7940 7950 7960 7970
 7980 7990 8000 8010 8020 8030 8040
 8050 8060 8070 8080 8090 8100 8110
 8120 8130 8140 8150 8160 8170 8180
 8190 8200 8210 8220 8230 8240 8250
 8260 8270 8280 8290 8300 8310 8320
 8330 8340 8350 8360 8370 8380 8390
 8400 8410 8420 8430 8440 8450 8460
 8470 8480 8490 8500 8510 8520 8530
 8540 8550 8560 8570 8580 8590 8600
 8610 8620 8630 8640 8650 8660 8670
 8680 8690 8700 8710 8720 8730 8740
 8750 8760 8770 8780 8790 8800 8810
 8820 8830 8840 8850 8860 8870 8880
 8890 8900 8910 8920 8930 8940 8950
 8960 8970 8980 8990 9000 9010 9020
 9030 9040 9050 9060 9070 9080 9090
 9100 9110 9120 9130 9140 9150 9160
 9170 9180 9190 9200 9210 9220 9230
 9240 9250 9260 9270 9280 9290 9300
 9310 9320 9330 9340 9350 9360 9370
 9380 9390 9400 9410 9420 9430 9440
 9450 9460 9470 9480 9490 9500 9510
 9520 9530 9540 9550 9560 9570 9580
 9590 9600 9610 9620 9630 9640 9650
 9660 9670 9680 9690 9700 9710 9720
 9730 9740 9750 9760 9770 9780 9790
 9800 9810 9820 9830 9840 9850 9860
 9870 9880 9890 9900 9910 9920 9930
 9940 9950 9960 9970 9980 9990 10000

FastDB - Fast Pairwise Comparison of Sequences

Release 5.4

Results file sq2.res made by on Tue 11 Apr 95 10:59:11-PDT.

```
Query sequence being compared: US-08-223-263-2 (1-1795)
Number of sequences searched: 358124
Number of scores above cutoff: 4841
```

Results of the initial comparison of US-08-223-263-2 (1-1795) with:

- Data bank : EMBL-NEW 2, all entries
- Data bank : GenBank 87, all entries
- Data bank : GenBank-NEW 2, all entries
- Data bank : UEMBL 41_87, all entries

1000000-	N
000000-	M
-	B
-	E
*	R
O	
F00000-	S
-	E50000-
-	Q
-	U
-	E
-	N
-	C
-	E
510000-	
-	
-	
-	
5000-	
-	
-	
-	
-	
1000-	
-	
-	
-	
500-	
-	
-	
-	
-	
100-	
-	
-	
50-	

SCORE	STDEV
0	8
199	399
598	798
997	1197
1396	1596
1795	

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
---------	------	--------	--------------------

33	33	13.50
33	33	

Times:	CPU	Total Elapsed
1	0.00	0.00
2	0.00	0.00
3	0.00	0.00
4	0.00	0.00
5	0.00	0.00
6	0.00	0.00
7	0.00	0.00
8	0.00	0.00
9	0.00	0.00
10	0.00	0.00
11	0.00	0.00
12	0.00	0.00
13	0.00	0.00
14	0.00	0.00
15	0.00	0.00
16	0.00	0.00
17	0.00	0.00
18	0.00	0.00
19	0.00	0.00
20	0.00	0.00
21	0.00	0.00
22	0.00	0.00
23	0.00	0.00
24	0.00	0.00
25	0.00	0.00
26	0.00	0.00
27	0.00	0.00
28	0.00	0.00
29	0.00	0.00
30	0.00	0.00
31	0.00	0.00
32	0.00	0.00
33	0.00	0.00
34	0.00	0.00
35	0.00	0.00
36	0.00	0.00
37	0.00	0.00
38	0.00	0.00
39	0.00	0.00
40	0.00	0.00
41	0.00	0.00
42	0.00	0.00
43	0.00	0.00
44	0.00	0.00
45	0.00	0.00
46	0.00	0.00
47	0.00	0.00
48	0.00	0.00
49	0.00	0.00
50	0.00	0.00
51	0.00	0.00
52	0.00	0.00
53	0.00	0.00
54	0.00	0.00
55	0.00	0.00
56	0.00	0.00
57	0.00	0.00
58	0.00	0.00
59	0.00	0.00
60	0.00	0.00
61	0.00	0.00
62	0.00	0.00
63	0.00	0.00
64	0.00	0.00
65	0.00	0.00
66	0.00	0.00
67	0.00	0.00
68	0.00	0.00
69	0.00	0.00
70	0.00	0.00
71	0.00	0.00
72	0.00	0.00
73	0.00	0.00
74	0.00	0.00
75	0.00	0.00
76	0.00	0.00
77	0.00	0.00
78	0.00	0.00
79	0.00	0.00
80	0.00	0.00
81	0.00	0.00
82	0.00	0.00
83	0.00	0.00
84	0.00	0.00
85	0.00	0.00
86	0.00	0.00
87	0.00	0.00
88	0.00	0.00
89	0.00	0.00
90	0.00	0.00
91	0.00	0.00
92	0.00	0.00
93	0.00	0.00
94	0.00	0.00
95	0.00	0.00
96	0.00	0.00
97	0.00	0.00
98	0.00	0.00
99	0.00	0.00
100	0.00	0.00

01:25:58.06 01:26:08.00

Number of residues: 307267536

Number of sequences searched: 358124

Number of scores above cutoff: 4841

Cut-off	raised	to 26
Cut-off	raised	to 30
Cut-off	raised	to 33
Cut-off	raised	to 39
Cut-off	raised	to 42
Cut-off	raised	to 44
Cut-off	raised	to 47
Cut-off	raised	to 49
Cut-off	raised	to 52
Cut-off	raised	to 55
Cut-off	raised	to 58
Cut-off	raised	to 63

Cut-off raised to 65.
Cut-off raised to 69.
Cut-off raised to 71.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

2 100% identical sequences to the query sequence were found:

Sequence Name	Description	Length	Score	Init.	Opt.	Sig.	Frame
1. HDMMLCMTL	Human c-mpl ligand (ML) mRNA,	1795	1795	1795	130.53	0	
2. HSMILCMTL	Human c-mpl ligand (ML) mRNA,	1795	1795	1795	130.53	0	

The list of other best scores is:

Sequence Name	Description	Length	Score	Init.	Opt.	Sig.	Frame
3. HS011025	**** 96 standard deviations above mean **** Human megakaryocyte growth an	1341	1338	1338	96.68	0	
4. HDMTA	**** 87 standard deviations above mean **** Human gene for thrombopoietin	7666	1216	1441	87.64	0	
5. HDMTHROMA	**** 83 standard deviations above mean **** Human thrombopoietin gene, co	6163	1156	1376	83.19	0	
6. HDMTHROMB	**** 76 standard deviations above mean **** Human thrombopoietin mRNA, co	1062	1062	1062	76.23	0	
7. MUSTHROA	**** 48 standard deviations above mean **** Mus musculus thrombopoietin m	1486	694	1155	48.97	0	
8. HSTNSCN	**** 10 standard deviations above mean **** Human mRNA for tenascin.	7286	180	780	10.89	0	
9. HDMHB	**** 9 standard deviations above mean **** Homo sapiens hexabrachion MRN	7390	180	781	10.89	0	
10. HSHXB	**** 9 standard deviations above mean **** Homo sapiens hexabrachion MRN	7390	180	781	10.89	0	
11. HSTENAS3	**** 9 standard deviations above mean **** Mouse ZP3 gene, encoding zona	1317	177	576	10.67	0	
12. MUSZP3	**** 9 standard deviations above mean **** Adenovirus type 5 fiber prote	2530	166	704	9.85	0	
13. ADRTIBA	**** 8 standard deviations above mean **** Discopryge omata Ca2+ channel	6981	164	701	9.70	0	
14. DTGCALB	**** 8 standard deviations above mean **** Murine GLUD mRNA for glutamat	2942	159	738	9.33	0	
15. MMGLUD	**** 8 standard deviations above mean **** G.gallus mRNA for integrin be	3482	154	770	8.96	0	
16. GINTB3	**** 8 standard deviations above mean **** Human mRNA fragment for plate	2129	153	362	8.89	0	
17. HSP1G2B1	**** 8 standard deviations above mean **** R.norvegicus beta-1-adrenergic	1645	152	724	8.82	0	
18. RATBIAR	**** 8 standard deviations above mean **** C.porcellus mRNA for preproac	1320	147	596	8.45	0	
19. CEPACRO	**** 8 standard deviations above mean **** X.laeviis POMC-B gene for proo	14776	144	697	8.22	0	
20. XIPOMCB	**** 8 standard deviations above mean **** Human cytochrome P45011B1 (et	14776	144	776	8.22	0	
21. HSCYITE	**** 8 standard deviations above mean **** C.eellipoides rbcL, rps14, tr	9705	143	735	8.15	0	
22. CHIC27	**** 8 standard deviations above mean **** C.eugametes l1637 gene.	2935	142	765	8.08	0	
23. CEL1637	**** 8 standard deviations above mean **** H.sapiens CAT mRNA.	2461	141	758	8.00	0	
24. HSCATMR	**** 8 standard deviations above mean **** Bacteriophage S13 circular DN	5386	141	486	8.00	0	
25. S13CG	**** 8 standard deviations above mean **** Human endocytelial nitric-oxid	23142	141	781	8.00	0	
26. HSGNOS48	**** 8 standard deviations above mean **** Rattus norvegicus interluekin	2598	140	519	7.93	0	
27. RAT12RBC	**** 8 standard deviations above mean **** Human RHD blood group antigen	1354	139	593	7.85	0	
28. HDMRDANTI	**** 8 standard deviations above mean **** H.sapiens mRNA for rhesus pol	1545	139	672	7.85	0	
29. HSRH11	**** 8 standard deviations above mean **** H.sapiens mRNA for rhesus pol	2790	139	747	7.85	0	
30. HSRH11T	**** 8 standard deviations above mean **** Yeast (S.douglasi) NAM2 gene	4072	139	648	7.85	0	
31. SDNAM2G	**** 8 standard deviations above mean **** Rattus norvegicus phospholipa	7203	139	761	7.85	0	
32. RATPHOSPHB	**** 8 standard deviations above mean **** T.therophilus gene for lyeyl	1738	137	694	7.70	0	
33. TLVYS	**** 8 standard deviations above mean ****						

34. CTE64CUD	C.coturnix clusterin gene.	6301	137	753	7.70	0
35. HSC1352	Human proto-oncogene (BCU3) g	4776	135	782	7.56	0
36. MDS19R	Mouse interluekin 9 receptor	3020	134	760	7.48	0
37. HDMENIA	Human mRNA.	3132	134	764	7.48	0
38. CHKA3A1	Gallus domesticus beta A3 cry	5327	134	724	7.48	0
39. HSP11V	Human pro-alpha-1 (V) collage	7138	134	781	7.48	0
40. HSWFAB	Human von Willebrand factor p	21033	134	756	7.48	0
41. ADMCOMPEN	Maedadenovirus h5 gene, compl	35935	134	776	7.48	0
42. HDMCYTMB	Homo sapiens (clone pAT 744)	640	133	247	7.41	0
43. HSCITNMB	Homo sapiens (clone pAT 744)	640	133	247	7.41	0
44. HDMPIIB	Human platelet glycoprotein I	703	133	304	7.41	0
45. HSGPIIB	Human platelet glycoprotein I	703	133	304	7.41	0

1. US-08-223-263-2 (1-1795)

HDMMLCMTL Human c-mpl ligand (ML) mRNA, complete cds.

LOCUS HDMMLCMTL 1795 bp ss-mRNA PRI 07-JAN-1995
DEFINITION Human c-mpl ligand (ML) mRNA, complete cds.
ACCESSION L33410
KEYWORDS c-mpl ligand.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
REFERENCE
1 (bases 1 to 1795)
de Sauvage, F.J., Haas, P.E., Spencer, S.D., Malloy, B.E., Gurney, A.L.,
Spencer, S.A., Darbonne, W.C., Henzel, W.J., Wong, S.C., Kuang, W.-O.,
Oles, K.J., Hultgren, B., Solberg, L.A.Jr., Goeddel, D.V. and
Eaton, D.L.
Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-mpl
ligand [see comments]
Nature 365 (6481), 533-538 (1994)
94261202
COMMENT
NCBI gi: 506826
location/Qualifiers
1..1795
/organism="Homo sapiens"
/sequenced_mol="cDNA to mRNA"
216..278
/gene="ML"
/codon_start=1
216..1277
/gene="ML"
/note="erythropoietin homology domain bp 66..522; NCBI
gi: 506827"

sig_peptide
sig_peptide

CDS

mat_peptide

BASE COUNT 420 a 586 c 382 g 407 t

[illegible][illegible]


```
|||||
AGAGCCAAATTCCTGCTGCTGTAACCAACCTCCAGCTCCCTGGACCAATTCGCCGATACCTAAGCAGG
120 730 740 750 760 770 780
970 980 990 1000 1010 1020 1030
ATACAGAACTCTTGAATGCACTGCTGAGCTCTTCCCTGGACCCCTCAGCAGAGACCCCTAGAGCCCGGAC
|||||
ATACAGAACTCTTGAATGCACTGCTGAGCTCTTCCCTGGACCCCTCAGCAGAGACCCCTAGAGCCCGGAC
790 800 810 820 830 840 850
1040 1050 1060 1070 1080 1090 1100
ATTTCCTGCAAGACATGACAGACAGAGCTCCCTCCAGCCCACTCCAGCTGATATTCCTTCCCAACC
|||||
ATTTCCTGCAAGACATGACAGACAGAGCTCCCTCCAGCCCACTCCAGCTGATATTCCTTCCCAACC
860 870 880 890 900 910 920
1110 1120 1130 1140 1150 1160 1170
CATCTCTCTCTGAGACAGATAGGCTCTTCCCTCCAGCCCACTCCAGCTGATATTCCTTCCCAACC
930 940 950 960 970 980 990
1180 1190 1200 1210 1220 1230 1240 1250
CCCCCTCTCTGACCCCTCTGCTCCAAAGCCCACTCCAGCCCTCTTAAACAGCCCTTACACCCAC
|||||
CCCCCTCTCTGACCCCTCTGCTCCAAAGCCCACTCCAGCCCTCTTAAACAGCCCTTACACCCAC
1000 1010 1020 1030 1040 1050 1060 1070
1260 1270 1280 1290 1300 1310 1320 1330
TCCCAAAATCTGCTCAGAGAGGTAAGGTTCTCAGACATCCGACATTCATGATGACATC
|||||
TCCCAAAATCTGCTCAGAGAGGTAAGGTTCTCAGACATCCGACATTCATGATGACATC
1080 1090 1100 1110 1120 1130 1140
1330 1340 1350 1360 1370 1380 1390
CCTTCCCTGAGAGGCGCCCTGAGAGACAACTGACAAATTTCTTCTCTCTGAAACCCCAAGCCCTG
|||||
CCTTCCCTGAGAGGCGCCCTGAGAGACAACTGACAAATTTCTTCTCTCTGAAACCCCAAGCCCTG
1150 1160 1170 1180 1190 1200 1210
1400 1410 1420 1430 1440 1450 1460
GTAAGAGGATACAGAGACTGAAAAAGGAAATCATTTTCACTGATTAATTAACCTTGAAGCTATTTT
|||||
GTAAGAGGATACAGAGACTGAAAAAGGAAATCATTTTCACTGATTAATTAACCTTGAAGCTATTTT
1220 1230 1240 1250 1260 1270 1280
1470 1480 1490 1500 1510 1520 1530
TTTAAGCTATACATTAATCACTGATTAATTAACCTTGAAGCTATTTTCACTGATTAATTAACCTTGAAGCTATTTT
|||||
TTTAAGCTATACATTAATCACTGATTAATTAACCTTGAAGCTATTTTCACTGATTAATTAACCTTGAAGCTATTTT
1290 1300 1310 1320 1330 1340
4. US-08-223-263-2 (1-1795)
HDMTA Human gene for thrombopoietin.
LOCUS HDMTA 7666 bp ds-DNA PRI 23-NOV-1994
DEFINITION Human gene for thrombopoietin.
ACCESSION D32046
KEYWORDS thrombopoietin.
```

```
SOURCE Homo sapiens blood DNA, clone lambdaHGT1.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
AUTHORS Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
1 (bases 1 to 7666)
Sohma,Y., Akahori,H., Seki,N., Horii,T.-A., Ogami,K., Kawamura,K.
and Miyazaki,H.
TITLE Molecular cloning and chromosomal localization of the human
thrombopoietin gene
JOURNAL FEBS Lett. 353, 57-61 (1994)
MEDLINE 95010765
COMMENT Submitted (08-Jul-1994) to DDBJ by:
Yoshiaki Soma
Pharmaceutical Research Laboratory
Kirin Brewery Co., Ltd.
2-2 Souja-machi 1 chome
Maebashi
Gunma 371
Japan
Phone: 0272-54-8618
Fax: 0272-52-2307.
FEATURES
source NCBI gi: 577319
location/Qualifiers
1..7666
/organism="Homo sapiens"
/sequenced_mol="DNA"
/tissue_type="blood"
<1420..1486
/number=1
/label=exon1
3157..3314
/number=2
/label=exon2
join(3302..3314,3546..3673,3960..4046,5979..6146,
6383..7048)
/note="NCBI gi: 577320"
/codon_start=1
/product="thrombopoietin"
/translation="MELTELILVVMILLIARLTSSPAPACDLRVLSKILRDSVLIH
SRISQCEVHRLPTPVLLPAVDLSLEMTQOEETRAODILGAVLLLEGMAARGOL
GPTCLISILQLSGOVALIGLALISLIGTOLPQGRTHAKDPNAILFISFHLIRGCV
RFILAVGSLICVBRAPPTAVSRISIVLTINELPNRTSGLETFNFSARTGSGL
LKMOGFRARIKIGLNOTSPSLDIDPGYINRIHEILNTRGLFPGPSRRITGAPDIS
GTSDTGSLPNIDPGISPSPTPTGQYTLFPLPTPLPVQLHPILDPSSAFPTPT
TSPLINTSYTHSONLSOEG"
3546..3673
/number=3
/label=exon3
3960..4046
/number=4
/label=exon4
5979..6146
/number=5
/label=exon5
6383..7576
/number=6
/label=exon6
1897 a 2154 c 1807 g 1808 t
BASE COUNT
ORIGIN
```


LOCUS	DEFINITION	ACCESSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
5. US-08-223-263-2 (1-1795)	Human thrombopoietin gene, complete cds.								
LOCUS	HUMTHROMA	6163 bp ds-DNA	PRI	29-JAN-1995					
DEFINITION	Human thrombopoietin gene, complete cds.								
ACCESSION	L36051								
KEYWORDS	thrombopoietin.								
SOURCE	Homo sapiens DNA.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria; Eutheria; Primates; Haplorhini; Catarrhini; Homnidae.								
AUTHORS	1 (baes 1 to 6163)								
TITLE	Poster,D.C., Sprecher,C.A., Grant,F.J., Kramer,J.M., Kuijper,J.M., Holly,R.D., Whitmore,T.E., Heipel,M.D., Bell,L.A.N., Ching,A.F., McGlane,V., Hart,C., O'Hara,P.J. and Lok,S.								
JOURNAL	Human thrombopoietin: gene structure, cDNA sequence, expression and chromosomal location								
	Proc. Natl. Acad. Sci. U.S.A. 91, 13023-13027 (1994)								

COMMENT	NCBI gi:	533214
FEATURES	Location/Qualifiers	
source	/organism="Homo sapiens"	
	/sequenced_mol="DNA"	
	<1..158	
exon	/number=1	
intron	159..1826	
	/number=1	
exon	1827..1984	
	/number=2	
CDS	join(1972..1984,2216..2343,2630..2716,4649..4816, 5053..5718)	
	/note="NCBI gi: 533215"	
	/codon_start=1	
	/product="tumorpoietin"	
	/translation="MELTELLLVMLLTARLTLSPPADCLRYLSKLRDHSVLT SRISQDEVHPPTPEVLLPADVFSIGEWKTOMETRAODILGAVTLLEGVAARGOL GPLTSLISGLSGGVRLILALOSLGTOIPGGRTAKDPNALFTSFQHLRGKV RFLMIQGSSTLCIVRAPPTTAVPSRTSIVLITNEIPRTSGILETFNFAARTGGVS IKMQGFRAKIRICGLINORTSRSLDOIPICTAIRIHLELNTRGLEPFGSRRTIAGDIS GRIDTSLPENLOPGISPSFTHPPGTITPLPPLTPVVQLHLPDPAFPPTP TSPSLNTSYHSNONLSORG"	
intron	1985..2215	
	/number=2	
exon	2216..2343	
	/number=3	
intron	2344..2629	
	/number=3	
exon	2630..2716	
	/number=4	
intron	2717..4648	
	/number=4	
exon	4649..4816	
	/number=5	
intron	4817..5052	
	/number=5	
exon	5053..>6163	
	/number=6	
BASE COUNT	1524 a 1706 c 1435 g 1498 t	
ORIGIN		
Initial Score = 1156 Optimized Score = 1376 Significance = 83.19		
Residue Identity = 81% Matches = 1436 Mismatches = 250		
Gaps = 86 Conservative Substitutions = 0		
X	10 20 30 40 50 60	
TCTTCCTAACCATCTGCTCCACAGAGGC-TGCCCTCCTGTGACATTGGGTCTCGAGCCCCTT		
CTGAACAACAGTGTAAGCACGGCTG-GACTTAGAGAGACAGGAGACTTTTAAACACAGAGCCTATGACCAG		
4420 4430 . 4440 4450 4460 4470 4480		
CT-----CAACCCGGATAG-ATTCTCTACCCCTTGGCCG-CCTTGGCCCCACCC--TACTTGGCCAGAG	70 80 90 100 110 120	
CTCTGGGAGACTGTGAAGAATTCCT-GCCCTGGGTGGGACCTTGTCTCTCCAGTTCACGCTGTATG		
4490 4500 4510 4520 4530 4540 4550		
-TGCAGAGCCTTAGCCGCTTC-ATGGC-CCACGAGAGGATTCAAGGGAGAGGCC---CAAACAGGAGAC	130 140 150 160 170 180	

[illegible]

11
ACAGACTTCGATTTGTTGGAGCAAACTTCACTCCCTACCGCAACTCTCTGGCTTGGCTTGAAGT
5320 5260 5270 5280 5290 5300 5310
880 890 900 910 920 930 940
GGCAGCAGGATTTTCAGAGCCAAAGTTCTCTGGTCGTCAACCAACCTCCAGGTCCTCGCAATCCCG
GGCAGCAGGATTTTCAGAGCCAAAGTTCTCTGGTCGTCAACCAACCTCCAGGTCCTCGCAATCCCG
5320 5330 5340 5350 5360 5370 5380
950 960 970 980 990 1000 1010 1020
GATCACTGAAACAGATACACAACTCTTGATGATGAACTCGTGAGCTTTTCTGAGCCCTCAGAGAGACCC
GATCACTGAAACAGATACACAACTCTTGATGATGAACTCGTGAGCTTTTCTGAGCCCTCAGAGAGACCC
5400 5410 5420 5430 5440 5450 5460
7530 1030 1040 1050 1060 1070 1080 1090
TAGAGAGCCCGGACATTTTCTCTAGAAACATACAGACAGAGCTCCCTCGACACCCAACTCCAGCTGGATTTT
TAGAGAGCCCGGACATTTTCTCTAGAAACATACAGACAGAGCTCCCTCGACACCCAACTCCAGCTGGATTTT
5470 5480 5490 5500 5510 5520 5530
1100 1110 1120 1130 1140 1150 1160
CTCTCTCCCAACCCATCTCTCTCACTAGATACAGTACAGCTTTCTCTCCATCCCACTTGGCCACCCCTG
CTCTCTCCCAACCCATCTCTCTCACTAGATACAGTACAGCTTTCTCTCCATCCCACTTGGCCACCCCTG
5540 5550 5560 5570 5580 5590 5600
1170 1180 1190 1200 1210 1220 1230
TGTGTACAGCTCACCCCGCTGCTCTTCTGACCCCTTCTCTCAAGCCGCAACCCCTACAGCCCTTCTTAAAC
TGTGTACAGCTCACCCCGCTGCTCTTCTGACCCCTTCTCTCAAGCCGCAACCCCTACAGCCCTTCTTAAAC
5610 5620 5630 5640 5650 5660 5670
1240 1250 1260 1270 1280 1290 1300
CATCTTACACCCATCTCCAGATCTGTCTCAGGAAGAGGTCTCTCAGCACTCGCCGACATCAACATTGT
CATCTTACACCCATCTCCAGATCTGTCTCAGGAAGAGGTCTCTCAGCACTCGCCGACATCAACATTGT
5680 5690 5700 5710 5720 5730 5740
1310 1320 1330 1340 1350 1360 1370 1380
CTCATGTAGAGCTCCCTTCCCTCTGACAGGCGCGCCCTGGGAGACAATGTGACAAAGATTCTCATCTTCTCTG
CTCATGTAGAGCTCCCTTCCCTCTGACAGGCGCGCCCTGGGAGACAATGTGACAAAGATTCTCATCTTCTCTG
5750 5760 5770 5780 5790 5800 5810 5820
1390 1400 1410 1420 1430 1440 1450
AACCACAAAGCCCGGTAAAGGAGATACACAGACTGAAAGGGAATCATTTTTCATCTGTACATTAATAACCT
AACCACAAAGCCCGGTAAAGGAGATACACAGACTGAAAGGGAATCATTTTTCATCTGTACATTAATAACCT
5830 5840 5850 5860 5870 5880 5890
1460 1470 1480 1490 1500 1510 1520
TCAGAACCTATTTTTTAAAGCTATACGAATACCTATACAGACAGCTAGCTTTTGGTGTATTTTCTGACAG
TCAGAACCTATTTTTTAAAGCTATACGAATACCTATACAGACAGCTAGCTTTTGGTGTATTTTCTGACAG
5900 5910 5920 5930 5940 5950 5960
1530 1540 1550 1560 1570 1580 1590
AATTGGCACTCACTATTTCTCTACATCTCTTTTCTGTGATTACTCTGAAAGGCGCTGGGCTGGCTGGC

530	540	550	560	570	580	590	ACCACTTGGCTGTCAATCCCTCTGGGGAGCTTTTGTGACAGAGTCCGTCTCTCTTGGGGCCCTGTGACAG
490	500	510	520	530	540	550	ACCTCTCGCTCTCAATCCCTCTGGGAGACAGCTTTCTGGGAGAGTTGCCCTCTTGGGGGGCCCTGTGACAG
600	610	620	630	640	650	660	CCTCTCTGGAACCCAGCTTCTCTCAACAGGCGAGACACAGCTCAACAAAGATCCCAATGCATCTTCTGTAG
560	570	580	590	600	610	620	CCCTCTAGAACCCAGCTTCTCTCAACAGGCGAGACACAGCTCAACAAAGATCCCAATGCATCTTCTGTAG
670	680	690	700	710	720	730	CTTCCACACACTGTCTCCGAGGAAAGTGCGCTTCTGTATGCTTTGTAGAGGGGTCAACCTCTGCGACGGCG
630	640	650	660	670	680	690	CTTGCACACACTGTCTTGGGGGAAAGTGCGCTTCTGTATGCTTTGTAGAGGGGTCAACCTCTGCGACGGCG
740	750	760	770	780	790	800	GGCCCCACCCACAGCTGTCCCAAGACAAACCTCTGTATGCTCTACACTGAACGAGCTCCAAACAGAG
700	710	720	730	740	750	760	GACCTCCCAACCAACAGCTGTCCCAAGAGACTTCTCACTCTCTCACTCAACAAAGTTCCTCAACAGAG
820	830	840	850	860	870	880	TTCTGGATTGTTGGAGCAAACTTCACTGTCTTCAGCCAACTACTGCTGTGGGCTTTGAAAGTGCAGCA
770	780	790	800	810	820	830	TTCTGGATTGTTGGAGCAAACTTCACTGTCTTCAGCCAACTACTGCTGTGGGCTTTGAAAGTGCAGCA
890	900	910	920	930	940	950	GGGATTTCAGAGCCAAAGAT---TCTGTGTCTGTGAACCAACTTCAGAGTCCCTGTGAACCAATCCCGGAT
850	860	870	880	890	900	910	GGGATTTCAGAGCAAGATTAATCTGTGTGTGCTAGTAATCAACTTCAGAGTCCCGGATTCGTGTGAAT
960	970	980	990	1000	1010	1020	CCTGAACAGATTAACAGAACTTGTATGATGACACTCTGTGACTTTTCTGTGAACCTTACGACAGAACCTT
920	930	940	950	960	970	980	CCTGAACAGAGCAACGGAGCTGTGAATGGAATCAATGAGGCTCTTGTGTGAACCTCACTTCACTTGA
1030	1040	1050	1060	1070	1080	1090	AGCCCCGAGCAATTCCTCTAGGAATGACAGACAGGCTCCCTCCACCAACCTTCAGAGCTGTGATTTCT
990	1000	1010	1020	1030	1040	1050	AGCCTTCAGCAATCTCGCCCGAGCTTTCACAAAGGCTCCCTGGGATTCAACTTCACAGGTTGAGCTTC
1100	1110	1120	1130	1140	1150	1160	TTTCCCAACCATCTCTCTATGTGACAGATTAAGCTTTCCCTTTCACACCACTTCCGACAC--CCTT
1060	1070	1080	1090	1100	1110	1120	TTTCTCAAGCTGTCTCTCTATGAGC---ACAGACCTTCCCTCTTCACTGCGCTTCCGACACCAAGAT
1170	1180	1190	1200	1210	1220	1230	-TC-----CAGTTCACACCCCTGCTTCTCTGTACACCTTCTGTCTCAACGGCCACCTTCACACCTT
1130	1140	1150	1160	1170	1180	1190	ATCTCACCCCAAGCTTCACACCCCTGTTTCTGTAGCCTTTCACACCAAGATGCTTAATCTTACCGGCT

[illegible]

[illegible]

LOCUS	HUMHXB	7390 bp ss-mRNA	PRI	08-NOV-1994
DEFINITION	Homo sapiens hexabrachion mRNA, complete cds.			
ACCESSION	M55618			
KEYWORDS	hexabrachion; tenascin.			
1	US-08-223-263-2 (1-1795)	Homo sapiens hexabrachion mRNA, complete cds.		
2	HUMHXB			
3	<pre> 1290 1300 1310 1320 1330 1340 TCAGACAC-TGCC-----GAC--ATCAGATTG-TCACA---TGTAAC-AGCTCCCTTCCCTCAGAGCGCC 2500 2510 2520 2530 2540 2550 --ACCA-CTGGCT--TGATTAACCTGGTTTCAAGCCCGCTGGTGAATGCATGCGATTGACCTACCTACGACA 2570 2580 2590 2600 2610 2620 2630 TCAAACACCTGTCAGAGAGACCGTACCCACCATGATCAGATCAAGAGAGACGAAGAACAT-ACCTCATCGGGAAC 2640 2650 2660 2670 2680 2690 2700 CT-GAAGCCTGACACTGATGTCAGAGGTCCCTC-ATCTCCCGCA--GAGTGACATGTCAAGCAACCGACAG 2710 2720 2730 2740 2750 2760 CAAGAGACCTTCACACAAGCGGCTCGATGCTCCACAGG-AT-----CTTC-GAGCTGTTCCAGACAGAG 2770 2780 2790 2800 2810 2820 2830 TATCAACAATACTCAATCAAGAGACGCTACCTCTTTC-CTCATTTTC--TGCAGAAATTGGACATCACTGA 2840 2850 2860 2870 2880 2890 2900 T-CTTGGAGGGGACACCGCTAGGTTGATTCCTCAAGAGCC--AACACGCC--ACA-----ACCAAAACCA 2910 2920 2930 2940 2950 2960 2970 CAGCTACAGGTCGAGCGCCGGAGATGCA-ATATGGAGATTGGAAGTTCTGCTGTGAAGAGAAACAAGAGAGC 2980 2990 3000 3010 3020 3030 3040 AACTGTGAGAAATGAATAGCTTTCTCTCAGAAAAAATTTTTTTTTTAAAAA 3050 3060 3070 3080 3090 AACTGACAGACCAAGCCCTGACCTGTGCT-AGAAAGACACGTTGGCCCAATTTCACCGCTAC </pre>			

SOURCE Human, cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria;
AUTHORS Eutheria; Primates; Haplorhini; Catarrhini; Hominoidea.
TITLE Nies, D.E., Hemeaath, T.J., Kim, J.H., Gulcher, J.R. and Stefansson, K.
The complete cDNA sequence of human hexabrachion (Tenascin). A
multidomain protein containing unique epidermal growth factor
repeats
J. Biol. Chem. 266 (5), 2818-2823 (1991)
MEDLINE 91131572
AUTHORS 2 (bases 1 to 7390)
TITLE Gulcher, J.R., Nies, D.E., Alexakos, M.J., Ravikant, N.A.,
MEDLINE Sturgill, M.E., Marton, L.S. and Stefansson, K.
92052108 Structure of the human hexabrachion (tenascin) gene
COMMENT Proc. Natl. Acad. Sci. U.S.A. 88 (21), 9438-9442 (1991)
FEATURES
NCBI g1: 184483
Location/Qualifiers
1..7390
/organism="Homo sapiens"
/sequenced_mol="cDNA to mRNA"
1..592
/partial
/gene="Hxb"
/number=1
/codon_start=1
136..6747
/gene="Hxb"
/map="9q32-q34"
/note="NCBI g1: 184484"
/codon_start=1
/gdb_xref="600-120-073"
/product="hexabrachion"
/translation="MGAMTQLAGVLAFLATLATEGCVLKKVIRHKRSGVNAIIPER
NOPVFNHVNINIKLPVSGQSVLDESAGSKDLAPPSESEFOEHTVDGNQVTFH
RINIPRACGCAAPDVKELISLELENIIVSLREOCTAGAGCCLOPATGRLDTPF
CSGRNFSTEGCGVCEGKMGKGNCSPEPCSGHLCRGICDGGCICDDGTGEDCSQ
LACPDNDGCKVNGVCICTFEGVAGANDCSREICPAPCEHGTGVDGICVCHGFAG
DDCKRPGLCNCRNRCRGVCEVCEGDEGTGEDCSLLICPNDGPRGRCINGTCCE
GTGEDCGKPTCPHACTGCGCEGCGVCEGAGLDCSEKCPADCHNRGRCVDRG
ECDDGFTGADGELKCPNGSGHRCVNGVCEGDEGTGEDCSQLCPNDCHNRGRC
EGKVCCEGFTGADGELKCPNGSGHRCVNGVCEGDEGTGEDCSQLCPNDCHNRGRC
GLAVDGGVCEGFTGADGELKCPNGSGHRCVNGVCEGDEGTGEDCSQLCPNDCHNRGRC
CHGGRCVCEGFTGADGELKCPNGSGHRCVNGVCEGDEGTGEDCSQLCPNDCHNRGRC
PPKDLVTEVTEETVNLAMNEMKRVETLYVYPTTEGGLBMOFRPMDQSTIIQEL
EPGEYFIRVAILLENKRSIPVSAVATYLPAGEKRSIKETISVEVEMDLDIAE
TWEIIFRNKKEDEGLITKSLRREPSTYRGTGLAPGOEYELSHIVKNTNPGIKRV
TTTSLDAPSOLEVDVDTTALITWFGPLAEIDGILTYGIDKVPDRTTIDITEDEN
OYSTGNKPDTEYVSLISRRGMSNPAKEFTTGLDAPRLNRVSDNDSITLEMK
NGKAAIDSYRIKVPISGSDHAEVDVCKSOATTTTIGIRPTEGXCIGVSAKXKX
ESNPATTAETLDPKDLQVSEYALSTLTKTLPALFDRRLNYSILPTGOMGVO
LPRNTTSVILGLEPGEQYVLLTAEGRRSKSPAHVASTEQAPLENLTVLEVMGV
GIRLMTADQAEHFTIIOEANKVEARNLTVPSLAVDIPGIKATPTVSYIG
VIOGRTPVLSAEASTGETPNLGEVVAEAGMDALKNMTAPAGAEYFTIOVODAT
VEAONLTVPGILSTDPGILKATHTTIRBGVODFTSTPLTSLVLELVEEPMGNL
TVTEVSMALRLNMTTDPGTDOFTIOVODAVEAHLNLTVPGLRSMEIPGIPACT
PYTVTLGEVGHSTRPLAVEVTEDELPQGLDIAVEVMDGIRLMTADQAEHFTI
IOVGEVKNVEAONLTVPGILRAVDIPGLEAATPVVSYIGVIRGTPVLSAEASTA
KEPEIGNLVSDITPESFNLSMAVDIGIFETFTIEIIDSNRLTEVYNI SCAERTAH

ISGLPSTDFIVYLSGLAPSIIRTKTISATVATLALPLENLTISDINPYGFTVSMAS
ENAPDSFIATVVOGSKLDPOEFTTSGTGRKLEGLITGIGYEWMLMIPRASKNP
LRAEIVTEAPEEDNDLVSDATPDGFRSLSWTDECVNPFYIKIRDKKOSPELEITL
LABERRDITGLKEATEYEIELXIGSKGRSQTVAIATTMAGSKREYISDITENSA
TVSWRAPTAOVESFRITVYITGGTSPVTVDTGTORLVKLLIPGEVYLSIIIAKKG
FEESEPVSGFTALDPSGLVYANTTDESLARWQPAIATVDSYVLSYGEKVEPIT
RTVSGNVEVALDIPATEYTLRIFAEKGPORSSTITAKFTTDDSPRLTAEVOS
ETALLTWPRPASVTGVLVYESVDGTVEKVIAGDPTTSYSLADSPSTHTAKIOL
NGPLRSMTIOTITFTTIGILYPPKDCOSAMINGDTTSGITTYIYINGDKAOLVFCDM
TSDGCGNIVFLRKNGRENTYONKAYAAFGDREPEFLMHLGIDNLAKITRAGQZEL
RVDLRDGHETAFVYDKEFVSGDKTRTKYKLVGYSGLADGSAVHNGRSFTDQKD
SAITNCALSYKGAFWYRNCRVNIMGRYDNNHSGOVNWFHWKGEHSIOFAEKRLRP
SNFENLEGRKRA"
593..2002
/gene="Hxb"
/number=2
/codon_start=1
2003..2266
/gene="Hxb"
/number=3
/codon_start=1
2267..2382
/gene="Hxb"
/number=4
/codon_start=1
2383..2539
/gene="Hxb"
/number=5
/codon_start=1
2540..2809
/gene="Hxb"
/number=6
/codon_start=1
2810..2995
/gene="Hxb"
/number=7
/codon_start=1
2996..3085
/gene="Hxb"
/number=8
/codon_start=1
3086..3349
/gene="Hxb"
/number=9
/codon_start=1
3350..3622
/gene="Hxb"
/number=10
/codon_start=1
3623..3895
/gene="Hxb"
/number=11
/codon_start=1
3896..4168
/gene="Hxb"
/number=12
/codon_start=1
4169..4441
/gene="Hxb"
/number=13

FT	exon	5923..6055	/gene="Hxb"
FT		/number=22	
FT		/codon_start=1	
FT		6056..6207	
FT	exon	/gene="Hxb"	
FT		/number=23	
FT		/codon_start=1	
FT		6208..6310	
FT	exon	/gene="Hxb"	
FT		/number=24	
FT		/codon_start=1	
FT	exon	6311..6469	
FT		/gene="Hxb"	
FT		/number=25	
FT		/codon_start=1	
FT	exon	6470..6636	
FT		/gene="Hxb"	
FT		/number=26	
FT		/codon_start=1	
FT	exon	6637..7390	
FT		/partial	
FT		/gene="Hxb"	
FT		/number=27	
FT		/codon_start=1	
XX			
SQ	Sequence	7390 BP; 1858 A; 1954 C; 2013 G; 1565 T; 0 other;	
Initial Score =	180	Optimized Score =	781
Residue Identity =	49%	Matches =	943
Gaps		Mismatches =	744
		Conservative Substitutions =	0

[illegible]

1640	1650	1660	1670	1680	1690	1700
ACTCCATCTGCTTGAAGCAAGACTGACGACCGACGAGGCTTCAACCTTGGCTTACACTGTC-CTG-CTT	340	350	360	370	380	390
CCGCCCTG-ACCTGCGAAGACTCT--CC--TGTCCAAAT-ATCTGCATTGGCC-----GGGGCTGTCTCT	1710	1720	1730	1740	1750	1760
GCCT--GCTGTGACTTTAGCTTG--GGA-GAATG-GAABAACCCAG-ATGAGAGACACCAAGCAACG-GAC	410	420	430	440	450	460
GAAATGGCAGCT-GCCTGTGCCATTAAGAGCAATTTTGGGCAAAAGCTGCAGAGAGCA--AAGATGTCCAGTGAAC	1770	1780	1790	1800	1810	1820
ATTGTGGAGACAGTACACCTTCTCTGCTGAGAG--ATG-ATGCGAG-CAGCGGAGC--AACTGGACACCAT	470	480	490	500	510	520
TGTGATG--CAGAGGCCGCTCTC--GTGAGAGGCGCAATGAT--CTGGCAGCAGGGCTTTCACAGCGCTGACTCTG	1840	1850	1860	1870	1880	1890
-TGCCTCTATCCCTCTCTGCGGCGACGCTTTCTG-GAAGGTCGC--TTCCTCTCTTGGGCGCCTCGAGAGCC	540	550	560	570	580	590
GTGC--CAGCACTTCTGCCCCAG-TGACTTGGCAAACTTGAACAATGTGCTTCGCGGCGCTGATCTGTG	1910	1920	1930	1940	1950	1960
TCCTTGGAAACCAAGCTTCTCTCAAGGCGAG-GACACAGCTCTCAACAAGATCCCAATGCCATTTCTTGAGG	600	610	620	630	640	650
AACAGAGGGCTTCAAGGAGGAAGACTGCTCTGACAGGTGTCTCTCCAAAGACCTGTTGTGACAGAGTGA-C-CA	1970	1980	1990	2000	2010	2020
TTTCAACACTGTCTCCAGGAAAGAGTGCCTTCTCTATGCTGTGAGAGGTCACACCTCTGCT-CAAGCG	670	680	690	700	710	720
GGAAAGACGCGTAACTTGGCGCTGGAACAT-GAGATGGGGTACAGAT--ACCTGT-CTGTACACG	2030	2040	2050	2060	2070	2080
GGCCCCACCAACAGAGCTGTCCAGAGCAACACTCTCTAATGCT-CAACACTGAACGAGTCC--CAATACAG	740	750	760	770	780	790
-----CCACCCACAGAGGGTGTGTGAAATGACTTCGTGTGTCTGGGAGCAACGCTCAACCATCATCAAG	2110	2120	2130	2140	2150	2160
GACTTTCGA--TTCT-TGGAG-AC--AAATCTACTCTCCACGACCAAGAACTAATGAGCTCTGGGCTTCTG	810	820	830	840	850	860
GA--GCTGGAACCTGTGTGGAGTACTTATTCGTGTAATTTGGCATTCCTGGAGAAACAAGACATTTCTTGT	2180	2190	2200	2210	2220	2230
-AAGTGCACACAGGATTACAGCCAAAGATTTCCGTGCTGTGGA--CCAAACCTCCAGAGTCC--CGAAC	880	890	900	910	920	930
TCACGCGCAAG--GGTGGCA--CGTACTTACCG--CACTTGAAGGCTGGAATTAAGTCATCAAGGA	2250	2260	2270	2280	2290	2300
CAAAATCCCGGATTAACCTGAACAGAAATACAGA-ACCTTGAATGAAGATCGTGA-CAT-TTTCCGAGCCCG	940	950	960	970	980	990
GAAATCTGTGGA-AGTGAATGGAGTCTCTTACATATGCTTTGAACCTGGAGATCATTTCCGGAAT	1000					

2310	2320	2330	2340	2350	2360	2370	2380
1010	1020	1030	1040	1050	1060	1070	
TCACGAGGACCCCT-AG---GAGCCCGGACATTTCTTAGAACAATACACAGAGCTCCCTGCACCCAA							
TCGAATAAAGAAATGAGGAGACAGATACCA-AAAAGCTAGG-AGCCAGAGACCTCTTACCGGAAATCGG							
2390	2400	2410	2420	2430	2440	2450	
1080	1090	1100	1110	1120	1130	1140	
CTCTCAGCCTGAGATTTCTCTCTCCCAACCCATCTCTCTACTG-AGACGATATAGCTCTTCCCTCTTCCAC							
TCCTAGCTCCCTGGGCAAGATATGATATCTC-TGCATCTAGTGAATAACATCTACCGGGGCCCTGGCTGTA							
2460	2470	2480	2490	2500	2510	2520	
1130	1160	1170	1180	1190	1200	1210	
CCACCTTGGCCACACCCCTGGTGGTCCAGCTCAACCCCTGCTCTCTTAACCTTTCTCTCAAGGCCACCCCTTA							
AGAGGGTGACCAACAC-ACGGCT-TGAGT--GCCCCAG--CCAGATGAGTGAATAAGATGCACAGAC-A							
2530	2540	2550	2560	2570	2580		
1220	1230	1240	1250	1260	1270	1280	
CCAACCCCTCTTCAAAACATCTCAC-ACCACCTCCCAAG-ATCTGTCTCAGGAAGGTA--AGTTTCTCA							
CCA-CTGCCT--TGATACCTCTGGTTCAGGCCCTGGCTGATGATCAGATTCGAGCTGACCTACCGCATCA							
2590	2600	2610	2620	2630	2640	2650	
1290	1300	1310	1320	1330	1340		
GACAC-TGCC---GAC---ATCACGATTTG-TCTCA---TGTAAC-AGCTCCCTTCCCTCGAGAGGGGCCCTCTG							
AAGACGTGGCAGAGACCCGTACACATCAATCAATCTCAAGAGACAGCAAGCAAGT-ACCTCATGGGAACCT-							
2660	2670	2680	2690	2700	2710	2720	
1350	1360	1370	1380	1390	1400		
GGAGAC---AACT--GGACAAATTTCTTAATTCTTCTCTGAAACCAAGCCCTGGTA--AAGGATATACAC							
GAAACCTGCACATGATAGAGAGTGTCCCTC-ATCTCCGCA--GAGTGCATGTCAAGCAACACCGACCAA							
2730	2740	2750	2760	2770	2780	2790	
1410	1420	1430	1440	1450	1460	1470	
AG-GAC--TGA-AAAGGAAATCTATTTCACTGACATTAATAACCTCAGACACTTTTTTTA-AGCTAAT							
AGACCTTACACAAAGGCTCGATGCTCCACAGA-TT-----CTTC-AGACGTGTTCCAGACAGATTA							
2800	2810	2820	2830	2840	2850		
1480	1490	1500	1510	1520	1530	1540	
CAGCAATACATCAAGAGAGCTAGCTTTTG-CTCATTTTC---TGGAGAAATTTGCAACATCATGATGTC							
CAGCATACCCCTGGAATGAGAGATAGCCAGGACAGCTATTGACAGTTACAG-ATTTAAGTATGCCCCACT-C							
2860	2870	2880	2890	2900	2910	2920	
1550	1560	1570	1580	1590	1600	1610	
TCATACATGCTCTTTTCTG-TGATATCTTCGCAAG-CCCTGGCGCTGGGACGTTGAACAGAGAGAGAG							
TCGTGAGGAGCAACGCGCTGAGTTTATGTTCCAAAGAGCC-AACAAGC--ACA--ACCAAAACACACAC							
2930	2940	2950	2960	2970	2980		
1620	1630	1640	1650	1660	1670		
ACTTAACCT-TGATCTCGAAAACAGAGAAAGGCTAATTTCTT--TGCT-TCA---AATTTCAAC---GCCTT							
TCACAGGCTGTAGGCGCGGGAAGCTG-AATATGGATTGGAGTTTCTGCTGTAGAGAAACACAGAGCAAT							

	1680	1690	1700	1710	1720	1730	1740	
CAGAGCCCCCACCCTTTACATCACTTCGATGGACTGTGAATCCCATATTTCTTAACAAGATCTT---A								
CCA--GCAGCACATCAACCCACA-CA-----GAGTGGACAC--GCCCAAGACCT-TCAAGTTTCTGAAA								
3070	3080		3090	3100	3110	3120		
1750	1760	1770	1780	1790	X			
CTTTTAGAAATGAATACGTTTCTCTCAGAAAAAAAAAAAAAAAAAAAAAAAA								
CTCAGACAGCAGCCCTGACCTGCTCT-GGAMAGACACCGCTGGCCAAATTGTGACCGCTACC								
3130	3140	3150	3160	3170	3180			

11. US-08-223-263-2 (1-1795)
HSTENAS3 H.sapiens mRNA for tenascin-C, 7560bp.

LOCUS	HSTENAS3	7560 bp	RNA	PRI	06-OCT-1994
DEFINITION	H.sapiens mRNA for tenascin-C, 7560bp.				
ACCESSION	X78565				
KEYWORDS	tenascin-C; wnaacin.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Mammalia;				
AUTHORS	Theria; Eutheria; Primates; Haplorhini; Catarrhini; Homnidae.				
TITLE	1 (bases 1 to 7560)				
JOURNAL	Cherzi,R., Carmella,B., Siri,A., Ponsasi,M. and Zardi,L.				
REFERENCE	Human Tenascin Gene: Structure of the 3' Region, Identification and				
AUTHORS	Characterization of its Promoter				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 7560)				
REFERENCE	Zardi, L.				
AUTHORS	Direct Submission				
TITLE	Submitted (26-JUL-1994) to the EMBL/GenBank/DBJ databases. Luciano				
JOURNAL	Zardi, Cell Biology Laboratory, Istituto Nazionale per la Ricerca				
FEATURES	sul Cancro, Viale Benedetto XV, 10, Genova, 16132, Italy				
source	NCBI gi: 556844				
	location/Qualifiers				
	1..7560				
	/organism="Homo sapiens"				
	/dev stage="Adult, fetal"				
	/tissue type="melanoma, fetal brain, breast, genomic from				
	placenta"				
	/sex="Male"				
	314..6919				
	/note="NCBI gi: 556845"				
	/codon start=1				
	/product="human tenascin-C"				
	/translation="MGAMTOLIAGVTFALATAGEGLKIVIRHRSGVNATLPER				
	NOPIVVNVIKILPVGSQSVLTESASGEKDLPSEPSEPOETWVGNIQITFTPT				
	RINIPRHAGCAAPDVEKLISREELIYNLSIRECTGAGCCLOPRTGIDTRBP				
	CSRGMPSTEGCCVCCEPGECGVNSPEPCGCHTGRIDICCDIGFGEDSCSL				
	LADSPNDQCKVNSVCICEFVGADCSNEICPPFCSEEHCTVDGICLVCHDFRA				
	DDNKRLCLNNCNRNRCVENECVDEBEFTDESILICPNDFDGRINCYCEE				
	GTEEDGCRPTCPHACHTNGREEGCVCGEFAGLDCSKRPADGNHRGCVGROR				
	ECDDGTGADCGELICPNCGSHGRVAGOCVEDGYTGDCSQLBCPDNCSRRCVCY				
	EGKGVCEOGFKGDCSDMSCPNDCHHGRCVNGKVCVDGTGEDCRDRCPDNRR				
	GLVDVQCICEDGFTGPCPAFLISCNDCHGRCVNGCVCHGCTMKDKCKQRPSD				
	CHEGRCVDGQCI CHEGTTGLD CGHS CP SDNN I AGCV SG RCI NE GTS GEDCSVS				

2990 3000 3010 3020 3030 3040 3050 3060

CDS

PRDLVTEVEETETVNLAMDNEMRTEYLVYVTPTEHSELEMOFNRPGDSTIIIOEL
ERCEVEFIKVPALIEENKKSIPVARSATYIAPPEGKEFKSEIKETSEVEEMPIIDIAE
TWEIIFRNANKDEGEETKSLRERETSRYOTGLABQCEYEISLHAKRNTSPGICLKV
TTTRIDAPSOILEVADVDTTALITWRPDLAEIDGLEITLIGKOVGDRTTIDLEDEN
QYSIGNIKPDEYEVSLSIRGDMSSNPAKEITFTGLDPRNLRLVRSQTDNSITLEWR
NGKAIDSVRIKAPISGDDHAEVDKRSQATYKTLGLRPEGLREYIGVSAWKEK
ENSPATINAEITDPTLOVSETAETSLITLKPPIAEFDRLNYSILPGVWVGVO
LPRMTSVYLRGLEPGEYVNLITAEKGRHSPAKVASTEOAPELNLNITVEVMD
GIRAMTRADQAEHFIIOVEANKVEARNLIVPSILAVDIPGLKATYTTYSYIG
VIQGRIPVLSAEASTETENLGEVVAEYGMWALMLNTAPEGAYEFTFOVEDADT
VEAONILVPGELSTDLPGIKATHYITIRKVTODFSTPLLSVETEEVEPMGNL
ITVEVMDALRLNMTTPDGTDOFTIOVEDADVEEHNILTVPGSIRSEVPEGLRAGT
PYTTLIGEVEHSTRPLAVEVETEDIPGLDIAVEYMDGIRLMTWALNNAEHEV
IOVEUNKVEAONILTPGSLRAVDIPGLEAPRYVTSYIGIRICRYRPIVSAESTW
KEPEIGNLVSDITPESFNLSMADTGEITTEIILIDNRLLEVEYNISGERTAH
ISGLPSTDFIVYLSGLAPSIKRTISATITAPLLENLTIIDINPISGTFVWMS
ENAFDSFLVTVVDSGLIDPOEFTLISGOKLELRLIIGIYEVWVGFTGQHQK
LRAELTEAEPEVDNLIVSDATPDGFLSWTADGVDNFVLKIDRTKOSPLEITL
LAPERTDITGLREATEYEIEIYGISKGRSOTVSAITTAGSREYVFSPTENSA
TVSWRAPAOVESPRITVYPIGTGTSMTVMDCTKTOTLXKILPGVVIYISITAMG
FESEPVSGSTTTLDGSPGLVANTIDSEALARMOPALATVDSIVYSTGKEPBIT
RTVSGNTEVALTDLEPAETYLIRIFAEKPKOSSTITAKFTTIDSDPDLATEVOS
ETALLTWPPRPAVSYGLIYVEVSDGVKVIYIGPTTYSIADLSPSHTAKIOL
NGPLRSMIOTITFTTIGLIPFEDCSOAMINDTISGLYITLINDKRAOLEVFCM
TSDGGWIVELRKNGRENFYOMWKAAYAAGFGBRREFFWIGDINLKIITAOQYELRY
DLDHGETAFVAVDKFSVGDAKTRKYKVEGSGTAGDSMAVHNRSPSTTDKODSA
ITNCALSTKCAFWRNCHRVNIMGRYDNNHSGVAMFWHMKHEHSIOFAPKELRPSN
FRNLEGRKRA*

polYA signal
BASE COUNT 1894 a 2019 c 2062 g 1585 t
7522..7527

Initial Score = 180 Optimized Score = 781 Significance = 10.89
Residue Identity = 49% Matches = 944 Mismatches = 740
Gaps = 232 Conservative Substitutions = 0

X
TCTTC-CTAACCAT--CTGCT--CCCCAGAGGGCTGCTGCTCATTTGGG-----TCCGCG
TGTCCCAATGCTGCTGAGTGGCCATGGCCCTGTGTCAATGGGCACTGTG-TGTGTATGAGGGCTATACGG
1540 X 1550 1560 1570 1580 1590 1600
60 70 80 90 100 110 120
AGCCCTTCTCCACCGGATA-GATTCTTCACCCCTTGGCCCGCTTGGCCACCACTTCTGCGCCAAAGTG
-GGAGAGCTGACCGACGCTACGGTGCCTCCCAATGACTGTCAAGTGGGGCGCTGTGTCAAGAGCAATGTG
1610 1620 1630 1640 1650 1660 1670
130 140 150 160 170 180 190
CAGAGCGCTAAGCGGCTCCATGGCCCCAGGA-AGGATTGAGGGAGAGGCCCA--ACAGGAG--CAC
TATGTAGCAAG--GCTTCAAGG--CTATGACTGAGTGAATGAGTGCCTTGAATGACTGTCAACAGAC
1680 1690 1700 1710 1720 1730 1740
200 210 220 230 240 250 260
GCCAGCAGACACCCCGGCAAGATGAGTGAATGCTCCTGCTGTCATGCTTCTCCATCTGAA
GCGCG-CTGTGTATGAGCATGTGTGTTGTGA-TG-AGGCTTACACAGGGAAGACTGCGCGGATGCCCA
1750 1760 1770 1780 1790 1800 1810

270 280 290 300 310 320
-GGCTAAGCTGTGCAAGC--CGGCTGCTCCTGCTGCTGCTCCAGCT--CTCCAGTA--ACTGCT--
TGCCCAAGGAGTGCAGCAACAGGGGCT-CTGTGTG-GAGGAC-AGTGCCTGTGAGAGCGGCTTACC
1820 1830 1840 1850 1860 1870 1880
330 340 350 360 370 380 390
---GCTGACT-CCCA---TGTCTTACAGCAGTGAAGCCAGTGGCCAGGTTACCTTGTGACAC
GCCCCTGACTGTGCAAGACTCTCTGTCCAAATAGC--GCCA-TGGCAG-GGT-CCGCTGTAAI-GGG
1890 1900 1910 1920 1930 1940
400 410 420 430 440 450 460
CTGCTGCTGCTGCTGTGCACTTTAGCTTGGGAGATGGAAMAACCAGATGAGAGCAAGGCACG-
CACTGCTGTGC--ATGAAGATTTA---TGGCAAA--GA---CTGACAGAGCA-AAGATGTCCAGT
1950 1960 1970 1980 1990 2000
470 480 490 500 510 520
GACATTCGTGAGAGTGTACCTCTGCTGAGAGG--AGTG-ANGCAG-CACGGGAC--AACTGGACCC
530 540 550 560 570 580 590
ACT-TGCTCTCATCCCTCTGCGGAGCTTCTG--GACAGTCCG---TCTCTCTTGGGCGCTGAGA
600 610 620 630 640 650 660
GCTCTCTGGAACCCAGCTTCTCCACAGGGGAG--GACACAGCTCCACAGATCCCAATGCCATTTCTG
TGCAAGAGGGCTACAGCGGAGAGACTGTGAGGTGTCTCTCCCAAGAGCTGTGTGAGAGAGTG
2130 2160 2170 2180 2190 2200 2210
670 680 690 700 710 720 730
AGCTTCCAAACACCTGCTCGAGGAAAGGTGCTTCTGATGCTTGAAGAGGCTCCACCTGTGGT-CAG
A-CGGAAGACGCTTAACCTGCTGAGCAAT--GAGATGCGGCTACAGAGT--ACCTGT-CTGTAC
2220 2230 2240 2250 2260 2270 2280
740 750 760 770 780 790 800
GCGGCGCCACCCACAGCAGCTGTCCAGCAGAAACCTCTTACTCT-CACCTGAACGAGCTTC--CAAA
ACG---CCACCCACAGGCTGTCTGAAATGCAATTCGCTGTGCGCTGGGAGACAGACGCTCACCATATC
2290 2300 2310 2320 2330 2340 2350
810 820 830 840 850 860
CAGACCTTCTGA---TGT-TGGAG-AC---AACTTCACTGCTCAGCAGCAAACTACTGCTGGCTT
CAGAA-gctGAGCCTGTGTGAGTACTTATTCCTGTAATTTTGCATCTCTGAGAAACAAGAGCATTC
2360 2370 2380 2390 2400 2410 2420
870 880 890 900 910 920 930
CTG-AACTGAGCAGAGATTAAGAGCAAGATTCCTGCTGTCTGAA--CCAACTCAGAGTTCCT---G
CTGACAGCGCAAG-gctGCCA---CTGACTTACTG--CACTGAAAGGCTTGAATTAATGCTCATAG
2430 2440 2450 2460 2470 2480

940	950	960	970	980	990	1000
GACCAATCCCGAATACCTAACAAGATACGA--ACTCTGAATGAGACTGCTGGA--CTC--TTTCCGGA						
GA--CACTCTGTGGA--AGTGGATGGATCTCTAGACATTGCTTTTAATTAACCTGGAGATATCTTCGGAG						
2490	2500	2510	2520	2530	2540	2550
1010	1020	1030	1040	1050	1060	1070
CCCTCAGCAGGAGCCCT--AG---GAGGCCCGGACATTTCCTTAGGAAATACAGACAGGCTCCCGCCAC						
ATATGATTAAGAAGATGAGGAGAGATACACA--AAAGCTCAGG--AGGCCAGAGACTCTTACGGGGAAAC						
2360	2370	2380	2390	2600	2610	2620
1080	1090	1100	1110	1120	1130	1140
CAACTCCACCCGTGAGATATTTCTCTCTCCCAACCCATCTCCCTACTG--GACAGTATAGCTCTTCCCTCTTCT						
TGCTCTAGCTCCCTGGGCAAGATGATGAGATATCTG--TGCAATATGAAAAACAATACCGGGGGCCCTGGCC						
2630	2640	2650	2660	2670	2680	2690
1150	1160	1170	1180	1190	1200	1210
CACCAACCTTGCCACACCCCTGTGGCTCAGCTCACACCCCTGGCTCTCAGCCCTCTCCACAGCCACACC						
TGAGAGAGGTGACACACAC--ACGGT--TGAT--GCCCCAG---CCAGATCAGAGTCAAGATATGCACAG						
2700	2710	2720	2730	2740	2750	2760
1220	1230	1240	1250	1260	1270	1280
CTACCAACCCCTCTTTTAAACATCTCAG--ACCACATCCACAG--ATCTGTCTCAGAAAGGTA---AGGTTCT						
---ACCA--CTGGCT--TGATCACCCTGTTCAAGCCCGCTGGGTAGATCATGAGTCAAGCTTACGCTGACCTACGCGCA						
2770	2780	2790	2800	2810	2820	
1290	1300	1310	1320	1330	1340	
TCAGACAC--TGCC-----GAC---ATCAGCATTG--TCTCA-----TGAC--AGCTCCCTTCCCTGACGGGCGCC						
TCAAAGCTGCCAGGACACCGCTACACATCATCATCACTCAGACAGCAGCAAGACAGT--ACTCCATGGGAGAC						
2830	2840	2850	2860	2870	2880	2890
1350	1360	1370	1380	1390	1400	
CTGGGACAC---AACT--GGACAAATTTCCCTACTTTCTCTCTGAACCAACCAAGCCCTGTAA--AAGGATAT						
CT--GAAACCTGACATGAGTACAGAGTGTCCCTC--ATCTCCGCA--GAGGTGACATGTCAAGCAACCCAGCA						
2900	2910	2920	2930	2940	2950	2960
1410	1420	1430	1440	1450	1460	1470
CACAG--GAC--TGA--AAAGGATCATTTTCTACTGTACATTAATTAACCTCAGAGCTATTTTTTA--AGC						
CAAGAGACCTTTCACACAGGCGCTCGATGCTCCACAGA--AT-----CTTC--GAGCTGTTTCCACAGACAG						
2970	2980	2990	3000	3010	3020	3030
1480	1490	1500	1510	1520	1530	1540
TATACGAAATATCTATACAGACAGCTACTCTTTTG--GTCTATTTTC---TGCAAAATTTGCACTCACTG						
TATACGATATACCTCTGATTTGAGATGAGATGGCAGAGCAGCTATTGACATTTACAG--AATTAGATATCCCGCCCA						
3040	3050	3060	3070	3080	3090	3100
1550	1560	1570	1580	1590	1600	1610
TTCTCTACATGCTCTTTTCTG--TGATATACCTTCCAAAG--GCCTGGGCTGGCTTGGGAGTTGACAGCAGG						
T--CTCTCGAGGACACAGCTGATGATGATCTTCCAAAGGCC--AAACAAGC--ACA---ACCAAAACCC						
3110	3120	3130	3140	3150	3160	

[illegible]

12. US-08-223-263-2 (1-1795)

MUS2P3 **Mouse ZP3 gene, encoding zona pellucida sperm-binding**

LOCUS	MUS2Z3	1317 bp	ss-mRNA	ROD	15-DEC-1988
DEFINITION	Mouse ZP3 gene, encoding zona pellucida sperm-binding protein, complete cds.				

KEYWORDS sperm-binding protein.

pZP3.2, and DNA, clone MZG I.

ORGANISM

Eukaryota; Animalia; Chordata; Vertebrata; Mammalia; Theria; Eutheria; Rodentia; Myomorpha; Muridae; Murinae.

REFERENCE

AUTHORS
Ringuette, M.J., Chamberlin, M.E., Baur, A.W., Sobleski, D.A. and Dean, J.

FILE

of the mouse zona pellucida
Dev. Biol. 127, 287-295 (1988)

MEDLINE

COMMENT Draft entry and computer-readable sequence [1] kindly submitted by J. Dean, 09-AUG-1988.

NCBI gi: 202462

FEATURES

Source

mRNA

sig_peptide

```

/note="ZP3 signal peptide (putative); putative"
/codon start=1
30..1304
CDS

```

```

/note=ZP3 precursor, NCBI gi: 202463*
/codon_start=1
/translation="MASSYFLICLLGCGELNLSQTLVLLFGGPTFVQSSSPKVKY
ECLEALIVATVSRLDFTGKGLVQEDLTLGEGEGQVSLVDITLVYAFNALHCESSV
OMTKDLATVSTFLDNPVNSGLSTLRNRRLVEYIECRQPEQVSSSHIIPOTVAPFR
ATVSSSEKLAFSLIRLMDENWTEKSAPTFLHGEVLAHLQVQGGSLPDLTVHCAVA
TPSLDPPNSPYHPIYDFHGLIVDGLSESSAQQVPEPDLITQTVDFVHFNSSSR
NTLYITLCHVAPANOIDPKLTKACSNFKNSQSMVLPQVPEPDLITQTVDFVHFNSSSS

```


mat_peptide

```

/note=sperm-binding protein (ZP3)
/codon_start=1

```

BASE COUNT	273 a	401 c	340 g	303 t
ORIGIN	Chromosome 6.			

Initial Score	=	177	Optimized Score	=	576	Significance	=	10.67
Residue Identity	=	49%	Matches	=	704	Mismatches	=	530
Gaps	=	196	Conservative Substitutions	=			=	0

180 X 190 200 210 220 230 240
 GCCCAACAGGAGGCACGCCAGCACAACCCTGGCCAGA--ATGAGGTACTTAATTGCTCTCTGG
 ||| ||| | | | | | | | | | | |
 CTGAG-----CGAGTGTAC-TCGAGCGGAGCATGGG-TCAGCTAATT--TCCTCTCC
 X 10 20 30 40 50

250 260 270 280 290 300 310
 TCATGCTTTCCCAACTGCAGAGCGTAACGCTGCA--GCCGGCGCTCCCTGCTGTGACACGAGGCTCT
 320 330 340 350 360 370 380
 CAGTAATCTGCTTGTGACTCCCATGTCTTCACAGACACTGAGCGCATGCGCAGAGGTCAACCTTTGCC
 390 400 410 420 430 440 450
 --GGA-ACTCC--CACCCAGAGGGGGCTCATTA-CGTGTGAAGTG--GAG--TGTCTGAAGC-
 120 130 140 150 160 170

[illegible]

460 470 480 490 500 510
CAAGCACAAGCATTTCTGGAGC--ACTGCCT-TCTGCTT--GAG-GAG--TATGG--CAG--CAC
|||||
CTTGCTCAGAAGGTTCACGCCCGGGTCCCTGTATCCACACCTGTGAGTTAAGCCCATTTGCAC

[illegible][illegible][illegible]

TAGANG-----GATCCACCCCTTGTCGTGAGCGAGGCCACCACACAGTGTGCCAGCAAACTCTCT
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
GAGAGAAACTGCCTTTCTCTCTTCGGCCTATGAGAGAACTGGAATA-CTG-----AG-AAATTCGGCTCC
530 540 550 560 570 580

780 790 800 810 820 830 840
 AACTCTACACTGACGAGCTCCCAACAGAGACTTCTGG-ATTGTTGGAGC---AAACTTCACTGCTCTGAG
 TTTCTTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
 CACTTTCACCTGGGAGAGGTATACCCAC---CTCAGAGCAAACTCAGACTGAGAAAGCCAC-CTGCCGCTGT
 590 600 610 620 630 640 650

[illegible]

920 930 940 950 960 970

ACCAAC-CTCAGGTCCTGGACCAATCCCGGATCC---TGACAGTATCAACAATT---GA
| | | | |
CCCATACTTCT--TC-GTGCAC---TTCCACGGTCCCTGTGGATGGCTCATCTGAAGGTTTTGGCAC
720 730 740 750 760 770 780

980 990 1000 1010 1020 1030 1040
 -TGAAC T-GTGA CTTTCTC T-GACCTACGACGAGCCAGAGCCGACATTTCTCAGAGAC
 TTTACAGTCCCGACGCCGACAGCTCTC---CAGTTACGGTGATGATTCATTTTCCACAGAC
 790 800 810 820 830 840 850

[illegible]

1120	1130	1140	1150	1160	1170	1180	1190
GACAGTATAGCCTCTTCCCTCTTCCACCCACCCCTTGCCACCCCTTGTCACGCTCCACCCCTGCTTCTCTT							
CTTCACAAAGCCTGTTTC--GTTCAACAGACTT--CCGACGAGTTCGTTTCCAG--TAGA--GGGTG--ATCTCTT	930	940	950	960	970	980	

	1200	1210	1220	1230	1240	1250	1260
ACCTTTG-C	CTCAA	GCCCA	CCCCCT	CTTTTAA	ACAATCTTA	ACCACATCCA	--GAT
ACATCTG	ATTCTC	ACCATG	GC--AC	TGTAGT	AATTCAG	CTTCAAGT	TCACATCCAGT
990	1000	1010	1020	1030	1040	1050	

1270	1280	1290	1300	1310	1320
TTCTCAGG--AAGGGTAGGTTTCACACACTCCG--ACATCAGC-AT-----TG-TCICA-TGTAACAGG					
CGCCAGTGGTTCACAGCT-AGTTTCTC-GAAACCCGACGACGACGATTCAGCTGATGTCTCACTGAGAGGGG					
1060	1070	1080	1090	1100	1120

TCC-----CTTCCCTGCAGGGGCCCCCTGGAGAGAACTGGACAAGATTTCCTACTTTCTCCTGAACCCAAA
 1330 1340 1350 1360 1370 1380
 ||| ||| ||| ||| ||| |||
 CCGCAATTCCTTGGAAAAAGCCACGACGAC--TGAGA--AGCGGTGAGATGC--TTTCGCTAAACC---
 1130 1140 1150 1160 1170 1180

1390	1400	1410	1420	1430	1440	1450
GGCCCTGGTAAAGGATACACAGAGCTGAAAGGAAATATTTTCACTGT-AAA-TTAAACCTTCAGAA						
-CTGTGGCTTGTGGGTT-----AGGCGTGGCAGACAGTGGCATTCCTGACCCCTGGCAGCATTAATAGCTTT-----						
1190	1200	1210	1220	1230	1240	1250
1460	1470	1480	1490	1500	1510	1520
GCTATTTTAAAGCTATACGACATACATCAACAGACACTGCTCTTTGGCTATTTTTCGCAAAATTG						
GC---TGTACACAGAAAGTGCACCTCTCTCTTCTACCTTGTATCCCTTCG-CAA-TAAAGAGAAATC						
1260	1270	1280	1290	1300	1310	X
CAACTACTG	1540					

13. US-08-223-263-2 (1-1795)
ADRFIBA Adenovirus type 5 fiber protein gene, complete cds

LOCUS	ADRF1BA	2530	bp ds-DNA	VRL	15-MAR-1989
DEFINITION	Adenovirus type 5 fiber protein gene, complete cds.				
REMARKS					

ANIMALS	caprine protein, rabbit protein
SOURCE	Adenovirus type 5 DNA, clone pECOR1 B
ORGANISM	Mastadenovirus h5

REFERENCE 1 (bases 1 to 2530)

TITLE The sequence of adenovirus fiber: Similarities and differences between serotypes 2 and 5
JOURNAL Virology 161, 549-554 (1987)

COMMENT	NCBI gi:	Location/Qualifiers
---------	----------	---------------------

```

/organism="Mastadenovirus h5"
476..2221
CDS

```

/trial=ation="MKRARPEEDTFNFNVYDYDETGPATVPTPLTPEFVSPNFGCESEPE
 GYLRISEPIVTSNGMALKMGNGLSLDAGNLTSQNTVTVSPPLKTKRKNINLEIS
 APLVITSZALIVYAAAPLMAAGNTLTQSOAPLTVMDSKLSIATQGPLVSEKALAL
 TSGPLTTDSSTLTITIASPPLVTNATGSLGIDILEYTNKGUGLKATPAHLVDFDNL
 TLTVATGCVIINNTSLQTKTGTALGDSQNNQINAVAGLQNSRNLILVDSYPTNL
 DADONOLRIAGGQPLFTINSANLIDINYNKGITFTSNNKSLEYLEITSTAGIEMEDAR
 AAINAAGCLEEGSPNAPNTAPLTKTGICHELPEFDSKAPMAGLGISIPDSGAIATVG
 MNKNDKLTLMWTPAPSEPCRLNAEKOKLTLVLTGKCSQILATVSYLAVKGSILAPIS
 TVQSOHLIIRDENGVILNNSFIDPEYWNRRNDLTGRIATYNAAGFENMISAPYKRS
 GKTASNNIVSOYLVINGDKTKTEVTLTITLINCQETGDTTSPASMSMDWSGNIYNEIN
 IATSSYFYSYLAOR"
 BASE COUNT 762 a 671 c 445 t 652 t
 ORIGIN 3 bp upstream of HpaI site

Initial Score	=	166	Optimized Score	=	704	Significance	=	9.85
Residue Identity	=	49%	Matches	=	874	Mismatches	=	646
Gaps	=	247	Conservative Substitutions	=			=	0

150 X 160 170 180 190 200 210
CATGGCCCAAGGAATTCAGGGAGAGCCCCCAACAGGAGCCACGCCAG-CCAGACACCCCGCCAG

X 10 20 30 40 50
GTAACTTCACCAGTG-----CAAA-AGGG-GTATCTTTGTCGTGTAAGCAGGCCA-A

[illegible]

280
GCCGCGCTCCCTCCT-TCGA-CCTCCAGTCTCAATAACGC-TTCG--TAGCTCCATGTCTTTCA
290
| | | | | | | | | | | | | | | | | |
300
AAATTGGTGTCATGTGGAGAAAAGCCATTAC-CATTACTCGACTCGGTAGAACAACGAAG-GCTGCC
310
120
130
140
150
160
170
180

330 360 370 380 390 400
 --GAGCAGCTACGCCAGTGGCCCAAGG-----TTCAACCTT-TGGCTTCACACCTGCTGCTG-CI--GCCGTGG
 TTCA-CTCACTGTTCAAGACCTTAGAATCTTCGACACCTTTATTAAGC-CTGGTGGCTTCCTCAAGAATCTC
 190 200 210 220 230 240 250

TTGTGACCTTAAAGCTGGGAGAAATGAAAAACCAAG-ATGAGAGAGACCAAGGCA--CAGGACATCTTGGGAGG
 TTTTCCCTTAACTAAATAAAAAATATATTAACACACTCTTCTTAATATAGTGGCAAAAT-TCTGTCCAA
 260 270 280 290 300 310 320

[illegible]

CCCTTCCTGGGAGCACTTTCGTGCACAGAGCTCCTCTCGTGAGGCCCTCTCTGAACACAGG
| | | | | | | | | | | | | | | | | | | |
AATTCTTCACAATCTTAATGCA-ATGTCACTTTCTCT---CTTCCGTGTCAATC-----GCACCCA-C-
400 410 420 430 440 450

[illegible]

690
CTCCGAGAAA--GGTCCGTTCATCATCCTTTGAGAGGGTCCACCTCTCGTCAGACGGGCCCAACACC
700
| | | | |
710
| | | | |
720
| | | | |
730
| | | | |
740
| | | | |
750
TACACGGAACCGGCCTCCACATGTCGCTTTCTTA--CTCTCCCTTTTATC-----CCCCA-----
530
540
550
560
570
580

CCACGACGCTCCGAGAGAACCTCCGAGGCTCCACGACGAGACGAGCCCAAC---AG-GACCTC---
 760 770 780 790 800 810
 ---ATGGCTTTAAGAG-AGTCCGCCCTGGGGATCTCTCTTGCGC-CTATCCGACCTCTATCTTCTCTCAAA
 590 600 610 620 630 640 650

820 830 840 850 860 870
GGATTCTTGAGACAACACT--TCA-CTGCCTCAGCCACAATA--CTGCG--TCTGGCCTTCTGAATGGA

```
GGCATCTTGGCGCTCAAAATGGGCAAGCGCCCTCTCTCTGAGCAGAGCGCGCAACCTTACCTCCCAAAATGTA
660      670      680      690      700      710      720
AGCA--GGGA--TTGAGAGC--C--AAGATTCTGCTGTCTGTAAC--CAAACTCCAGGTCCCTGAGCAAAATCC
880      890      900      910      920      930      940
ACACGTGTAGCGCCAGCCTCTCAAAAACCAAGTGC--AAACATAAACCTGGAAATATCTGACAC--CCTC-
730      740      750      760      770      780      790
950      960      970      980      990      1000
CCGATACCTTGAACAGATATACAGAACTCTTGG--AATGAACTGCTGAGCTCTTCTCTG--ACCTTCAC
1010      1020      1030      1040      1050      1060      1070      1080
GCAG--AACCCTAGAGAGCCCGGAGATTTCTCAGAGAACATCAGACAGAGCTCCCTGCGCAACCTTCA
1090      1100      1110      1120      1130      1140      1150
ACAGATTACCT--CAG--AAGCCCTAAGTGTGCTGCCCGCCACCTCTAATGTGGCGGGGAGCAACATCAG
860      870      880      890      900      910      920
GCTGTGATTTCTCTTCCCAACCCATCTCTCTACAGAGATATAGCTTCTCCCTCTTCAACAC--
1090      1100      1110      1120      1130      1140      1150
--CCT--CACAGTTCAGAGAAAGCTAGCC-----CTGCAAAACATCAGGC-----CCCTGC--ACCA--CGACCGA
930      940      950      960      970      980
TTGC--CCACCCCTGTGTGTC--CAGCTCCACCCCTGTCT--TCTGTACCCCTTCT--GCTCCAGCCGCAACCCCT
1160      1170      1180      1190      1200      1210
TACAGTACCTTACTATACCTACCTGCGCCCTCT--CTAATCTACCTGCTGCTGTGGCAATGACTTGA
990      1000      1010      1020      1030      1040      1050
1220      1230      1240      1250      1260      1270      1280
ACAGCCCTCTTCTTAAACATCTTACACACCTCCAGAAATCTGTCTCAGAGAGGTTAGGTTCTCA--GAC
1290      1300      1310      1320      1330      1340      1350
AAGGCCCATTTTATACACAAATGAAA---ACT---AGGA--CT---AAAGTACGGGGCTCTCTGATGTA
1060      1070      1080      1090      1100      1110
1290      1300      1310      1320      1330      1340      1350
ACTGCGCAGATTCAGCA--TTG--TCTCATGTACAGCTCC---CTTCC--CTCCAGGGGCG--CCCTGGAGAGC-
1360      1370      1380      1390      1400      1410      1420
ACAGACCTTAACACTTTGACCCGTAAGCACTGGCTCAGGTGATATTAATATATCTTCTCTGCAAACT
1120      1130      1140      1150      1160      1170      1180
1360      1370      1380      1390      1400      1410      1420
AACTGGCAAGATTTCTTCTCTCTGTAACCAAGAGCCCTGTGTAAGAGATAC--CAGGACTGAAAAG
1430      1440      1450      1460      1470      1480
AAGTTACTGGA--GCCT--TGGGTTTATTCACAGAGCAATATGCACTTAATATGAGCAGGA--GCACTAA
1190      1200      1210      1220      1230      1240      1250
1430      1440      1450      1460      1470      1480
GGAATCATTTT--TCACTG--TACATTATTAACCTTCA--GAAGCAT--TTTTTTAAGCT--ATCAGCA--TA-
1490      1500      1510      1520      1530      1540
GATGTTCTTCAAAACAGAGCCCTTATA---CTTGATGTATGATATCCGTTGATGCTCAAAACCACTAA
1260      1270      1280      1290      1300      1310      1320
--CT--CATAGAGCAGTACGCTTTTGGTCTAATTTTCTGAGA--AATTTG-----CACT--CA--CTGA--TT
1490      1500      1510      1520      1530      1540
```

```
ATTAAGACTAGAGAGAGGCGCCCTCTT--TTATTAACAGCCACACTTGTGATTTAATTAACAAAGGC
1330      1340      1350      1360      1370      1380      1390
1550      1560      1570      1580      1590      1600      1610
CTCTACATGCTCTTTTCTG--TGATTAACCTTGC--AAAGGCTTG--GGCTGGCCT--GGCAGTTTAAACAGAGG
1620      1630      1640      1650      1660      1670
CTTACTTCTG---TTTACAGCTTCAAAATTCACAAAAGGTTGAGTTAGCTTAACAGCA--CTGCCAAAGGGGT
1400      1410      1420      1430      1440      1450      1460
AGA--GACTAAC--CTTGAGTCAAAAACAGAGAAAG--GTTAATTTCTTGTCTTCAAAATTCAGAGCTTC--C
1620      1630      1640      1650      1660      1670
TGATGTTTACCTTACAGCA--TACCA--TTAATGACAGAGATGGGC--TTGAATTTGGTTCACCTTAATGCAC
1470      1480      1490      1500      1510      1520
1680      1690      1700      1710      1720      1730      1740
AAGCCCCCATCCCT--TTATATATATTTCTCAGAGGAGCTGTATCCATTTTAAACAGATCT--TTACT
1750      1760      1770      1780      1790      1800
CAAAACAATCCCTCAAAAACAAAATTTGGCAT--GGCTAGAAAT--TTGATTC--AAACAAGCTATGCTT
1530      1540      1550      1560      1570      1580      1590
1750      1760      1770      1780      1790      1800      1810
CTTGAGAAATGAATTAAGCTTTCTCTCAGAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTT
1820      1830      1840      1850      1860      1870      1880
CCTAAACTAGGAACTGGCTTGTGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
1600      1610      1620      1630      1640      1650
14. US-08-223-263-2 (1-1795)
DYGCCA1B      Discopyge ommata Ca2+ channel alpha 1 subunit gene
LOCUS      DYGCCA1B      6981 bp ds-DNA      VRT      22-JUL-1993
DEFINITION      Discopyge ommata Ca2+ channel alpha 1 subunit gene sequence.
ACCESSION      L12532
KEYWORDS      calcium channel alpha-1 subunit.
SOURCE      Discopyge ommata (library: lambda ZAP) adult electric lobe, brain
DNA.
ORGANISM      Discopyge ommata
Eukaryota; Animalia; Chordata; Vertebrata; Chondrichthyes;
Elasmobranchii; Euselachii; Neoselachii.
REFERENCE      Horne,W.A., Ellinor,P.T., Inman,I., Zhou,M., Tsien,R.W. and
AUTHORS      Schwarz,T.L.
TITLE      Molecular diversity of Ca2+ channel alpha 1 subunits from the
JOURNAL      marine ray Discopyge ommata
MEDLINE      Proc. Natl. Acad. Sci. U.S.A. 90, 3787-3791 (1993)
COMMENT      93248175
FEATURES      NCBI gi: 290386
source      Location/Qualifiers
1..6981
/organism="Discopyge ommata"
/cell_type="neuron"
/dev_stage="adult"
/sequenced_mol="DNA"
/tissue_type="electric lobe, brain"
/tissue_lib="lambda ZAP"
BASE COUNT      1858 a 1670 c 1767 g 1686 t
ORIGIN
Initial Score = 164 Optimized Score = 701 Significance = 9.70
```


CCAAAGCCAGAGCTATGAGAGAAAGCATCTTGGAGGCTGACTCTGATCATTTGATTCT-----GCTGGCA	1170	1180	1190	1200	1210	1220
330	340	350	360	370	380	
TGCTAGCTCCCATGCTCC-----TTACAGACACATGAGACATGCGCCAG-----GGTTCA-----CCCTTGGCC	1170	1180	1190	1200	1210	1220
1170	1180	1190	1200	1210	1220	
CGGAGAACATTTATACCAATTCATATGACACCGAGAGTCAAAGTCAGCAAGATCATTTGCTGAAGTGGCAATGGGC	1230	1240	1250	1260	1270	1280
1230	1240	1250	1260	1270	1280	1290
390	400	410	420	430	440	450
TACACCTTCCTCGCTGCTGCTGTGGACTTTAGCTTGGAGATGAGAAACCAACAGATGGAGAGAC--AAGG	1300	1310	1320	1330	1340	1350
CA--ACCACTCCAG--AGGCTGAT--AAGTTT--CGT--GAAAGAAATC--ATGGTATTCCGATCT	1300	1310	1320	1330	1340	1350
460	470	480	490	500	510	520
CACAGACATTTCTGAGAGACATCCACCT--TC--TGC--GAGGAGAGTGATG--GCAGCACGGGAGCAATGGG	1360	1370	1380	1390	1400	1410
CTACTTAATGCT--GAGAGGATTAACATGTTCTTACTTGAAGTGGCTAAAGATCTAAATCATGTCAGTAGG	1360	1370	1380	1390	1400	1410
530	540	550	560	570	580	590
ACCACCTTGCCTTCATCCCTCCCTGGGGGAGCACTTCT--GGACAGTGCCTCTC--CTCCTTGGGGCCCTGACAG	1440	1450	1460	1470	1480	1490
GCCCA--TTGACCTTCAATATATGAAAGGACTCTTACCTACCACTTCGTCTATGTCGTTCAGAGACTTTAGAG	1440	1450	1460	1470	1480	1490
600	610	620	630	640	650	660
AGCCTCTTGGAAACCCA--GCTTCCCTCCACAGAGGAGACACAGTCACAGAGATCCCATATCCATCT--TCCC	1510	1520	1530	1540	1550	1560
AGAAAGTTGGAAAGCATGTTGGAAATCTATTCGTGTGGTCCCA--CAGCAGTTTCCAGCAGAGATATCG	1510	1520	1530	1540	1550	1560
670	680	690	700	710	720	730
TGAGCTTCCCAACACCTGCTCCGAGGAAAGGTCGTTTCTGATGCTTGTAGAGGGGTCCACCCCTCTGGCTCA	1580	1590	1600	1610	1620	1630
GGTGATCTGGAAAGCATGTTGTCACCTGCGCTGGCCCTACACATGGA--GAG--ATCGCCAGGCAATTA	1580	1590	1600	1610	1620	1630
740	750	760	770	780	790	800
GGCGGGCCCCACCCACACAGCTCTCCCAACA--GAACTCTCTGTGCTCTCACTGAACGAGCTCCAAAA	1650	1660	1670	1680	1690	1700
TGCC---CAGAGCCATGA--ATTATTAACCTGGAGTTGACCTCGAGACAGTGGCTTATGTCATGCTATACGAG	1650	1660	1670	1680	1690	1700
810	820	830	840	850	860	870
CAGGACTTCTGGATGTTGGAGACAAATTCACAG--CCT--CAGCCAGAACTACTGGCTCTGGCTTTGAAAG	1710	1720	1730	1740	1750	1760
AAAGTCTCAAG--GTGT--ACATATGAGCTGTGTGACCTTCATCAATG--ACAGTTCACAGCCACATCTT--T--T	1710	1720	1730	1740	1750	1760
880	890	900	910	920	930	940
TGGCAGCAGGAGATTCAGAGCCAAATTCCTGCTGTCTGTGAACCAAACTCCAGG--CCCTGGACAAATC--	1780	1790	1800	1810	1820	1830
TACACG---TCTTC--ACATATATTTCTTCAGACCTGTGCACAAAGTTTCAATATATACATAG--AAATCT	1780	1790	1800	1810	1820	1830
950	960	970	980	990	1000	
-CCGGAATACCTGAACAGATACAGCAACTCT-----TGAATGGAATC--GT--GGACTC--TTTCTCGAACCTT	1780	1790	1800	1810	1820	1830
1780	1790	1800	1810	1820	1830	

GTCTCTGTGATCTCATTA--GTTA--AOGGACACTGTCTGAA--CAAGTCACTGTGAATACAGCCCTTAAGAA	1840	1850	1860	1870	1880	1890	1900
1010	1020	1030	1040	1050	1060	1070	
CAGCAGAGACCCTAG--GAGCCCCGAGCATTTTCCCTA--GGAAATCAGACAGG--CTCCCTGACAGCCACCC	1910	1920	1930	1940	1950	1960	1970
AAGAGATTAGTAAAGGATCATGTACAA--AGTGAATGATTAAG---TAGAATCACTGTATCCAA---	1980	1090	1100	1110	1120	1130	1140
TCGACGCTGATATTTCTCTCCCAACCAACCATCTCTCACTGGAACAGTATAGCTCTTCCCTTCCACCCA	1990	2000	2010	2020	2030	2040	
GCCAG--TTAGTGTGTTGCTTTTAAAGTAACTGTCTGCTCTGTCTCT--TGCCTGAGGCCCTTCCAGCA	1150	1160	1170	1180	1190	1200	1210
-----CTTGC-----CAACCCCTGTGGCTAGCTCCACCCCT--GCTTCCTGACCTTCTGTCACAGGCC	1220	1230	1240	1250	1260	1270	1280
GAGTCATGCTGTGGCAGAGAGAGCT--CAGAGAGACATGACTCTTGAATATCTG--GGAACTCCGG	1290	1300	1310	1320	1330	1340	
ACCCCTACAGCCCTTTCTTAAACATACCTCAACACCACTCCACAAATCTGTC---TCAGAAAGGTAAGGT	1350	1360	1370	1380	1390	1400	1410
ACAAATTAAGATGAGTGGTGTGGCTTTTCCA--AAGTGGCATGTCTC--CTCAATGATATTTGTGTTT	1420	1430	1440	1450	1460	1470	1480
TC--TCAGACACTGCCGACATGACGATGCTCATGTAC--AGCTCCCTTCCTCAGAGGCGCCCTGGAGAGA	1490	1500	1510	1520	1530	1540	1550
TCATTCAAAAAATAATTAAC--TCAG--TTTCTTAAGTTGCAACAAAAATAAAGCTATG--TTCAT--GATTTT	1560	1570	1580	1590	1600	1610	1620
CTTTTCTGTGA--TAATCTGCAAAAGCCCTGG--CTGGCTCGGACGTGTAACAAGAGAGACATTA--C--CA	1630	1640	1650	1660	1670	1680	1690
C--AGACTCGAAGGACAGG--GACTCTTGTGAACTGGTAATGACGT---CCAGAGAG--TAACGTAAACAAG	1700	1710	1720	1730	1740	1750	1760
TTGAGCTAGAAACAGAAAGGATTAATTTCC--TTTG--CTTTAAT---TCAAGGCGTT---CMAACCCGCC	1770	1780	1790	1800	1810	1820	1830

maryh@stic

stdin

NeWSprinter20

Tue Apr 11 13:31:59 1995

NeWSprint 2.5 Rev B

Openwin library 3

NeWSprint interpreter 210.0

NeWSprint 2.5